

# Essays in optimal fiscal policy



**Jan Kvasnička**

Faculty of Economics  
University of Cambridge

This dissertation is submitted for the degree of  
*Doctor of Philosophy*

Darwin College

September 2017



This thesis is dedicated to my parents.



## **Declaration**

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

Jan Kvasnička  
September 2017



## Acknowledgements

First and foremost, I would like to thank to my Supervisor Elisa Faraglia for her patient guidance and extraordinary level of support in all aspects of my PhD studies. I am also very grateful to my Research Advisor Charles Brendon and to Libor Dušek for their invaluable advice and support on the job market. I am thankful for helpful comments to Albert Marcet, Pierre Yared, and various participants of macroeconomics workshops at University of Cambridge. Chapter 4 (*Open economy optimal fiscal policy*) is the outcome of joint work with Elisa Faraglia and Rigas Oikonomou.

I gratefully acknowledge financial support of the Economic and Social Research Council and of the Faculty of Economics Trust Funds.

This work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (<http://www.hpc.cam.ac.uk/>), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council.

On a more personal note, I would like to thank to my parents and grandparents. Their love, investment in human capital, and continuing support were instrumental in allowing me to write this thesis. Very special thanks go to Marion for giving me a reason to smile every day, sticking with me through thick and thin, and taking good care of me when I was writing up.

I am also grateful to my friends both in Cambridge and elsewhere for all the good times, and keeping my spirits up in challenging times. In particular, I would like to mention my oldest friend Ondřej, and the many friends I made at Darwin College Boat Club.

Big thanks go to Emile for letting me use his room in the last few months of writing this thesis, and to Meltem and Michael for being absolutely wonderful housemates.





**Author:** Jan Kvasnička

**Title:** Essays in optimal fiscal policy

## **Abstract**

This thesis is of the three article format. All three articles contribute to the literature on optimal fiscal policy with exogenous government expenditures and distortionary taxation following Lucas and Stokey (1983) and Aiyagari et al. (2002) (AMSS).

The first article extends the framework of AMSS by modelling agents ex ante heterogeneous in deterministic labour productivity trends in an infinite-horizon production economy with incomplete markets. The government does not use transfers. When the productivities of different agents grow at different rates, there is a conflict over the timing of tax collection. This is explored in a two-period model. The infinite-horizon model with two agents ('low-skilled' and 'high-skilled') is used to quantitatively analyse the impact of productivity trends observed in recent decades on the optimal policy. The impact is significant. The model can contribute to explaining the increase in government debt in many advanced economies in recent decades. The optimal policy strongly depends on Pareto weights but welfare of the agents does not. Political economics implications are discussed.

The second article analyses the impact of heterogeneous productivity trends on the optimal policy when the social planner can use transfers. There is now conflict over the timing and the level of taxation, and it is explored in a two-period model. The optimal policy is studied in the same environment as in the first article. For most Pareto weights, the change in the tax rate is less pronounced than in the model without transfers, but still greater than the expected change due to shocks. The optimal policy and the welfare of the agents strongly depend on Pareto weights. Policy implications are discussed. The optimal policy in the horizon of decades is significantly affected by even a modest heterogeneity in the growth rates of the agents. Solution methods common to all three articles are discussed.

In the third article the closed economy model of AMSS is extended into an open economy setting with two countries. The government of each country finances its exogenous stochastic expenditures by distortionary labour taxation, and issues one-period bonds. The Ramsey planner chooses policy for both countries, and a no-arbitrage condition on the return of bonds of the two countries restricts her choices. The optimal policy is quantitatively studied in a calibrated model with ex-ante identical countries and equal Pareto weights, and three settings are compared in terms of policy and welfare: autarky (closed economy), partial union (international borrowing allowed), and full union (transfers between governments allowed).



# Table of contents

<b>List of figures</b>	<b>xv</b>
<b>List of tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Heterogeneous labour productivity trends and optimal fiscal policy</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Two-period model . . . . .	14
2.2.1 The individually optimal policy . . . . .	16
2.2.2 Limitations of the two-period model . . . . .	20
2.3 The full model . . . . .	21
2.3.1 The environment . . . . .	21
2.3.2 Competitive equilibrium . . . . .	24
2.3.3 The Ramsey problem . . . . .	26
2.3.4 Recursive formulation of the Ramsey problem . . . . .	27
2.3.5 Solution methods . . . . .	29
2.4 Increasing inequality and optimal fiscal policy . . . . .	32
2.4.1 Calibration . . . . .	32
2.4.2 Results . . . . .	35
2.5 Political economics implications . . . . .	41
2.6 Conclusion . . . . .	44
<b>3 Heterogeneous labour productivity trends and optimal fiscal policy with transfers</b>	<b>45</b>
3.1 Introduction . . . . .	45
3.2 Two-period model of optimal policy with transfers . . . . .	48
3.3 The full model . . . . .	54
3.3.1 Competitive equilibrium and the optimal policy . . . . .	56

3.4	Quantitative analysis of the optimal policy . . . . .	58
3.4.1	Increase in earnings inequality since the 1980s . . . . .	58
3.4.2	Policy implications of future productivity trends . . . . .	65
3.5	Solution methods . . . . .	67
3.5.1	Overview of solution methods . . . . .	67
3.5.2	Discrete approximation of the value function . . . . .	69
3.5.3	Parametric approximation of the value function . . . . .	70
3.6	Conclusion . . . . .	72
<b>4</b>	<b>Open economy optimal fiscal policy</b>	<b>75</b>
4.1	Introduction . . . . .	75
4.2	A two-country model of fiscal policy . . . . .	82
4.2.1	Competitive equilibrium and the Ramsey plan . . . . .	84
4.2.2	Full union . . . . .	89
4.2.3	Autarky . . . . .	91
4.2.4	The implications of the no-arbitrage constraint . . . . .	92
4.3	The Ramsey problem . . . . .	93
4.3.1	The Ramsey problem (sequential version) . . . . .	93
4.3.2	The Ramsey problem (recursive formulation) . . . . .	94
4.3.3	Solution methods . . . . .	96
4.4	Numerical examples . . . . .	97
4.4.1	Calibration . . . . .	98
4.4.2	Results . . . . .	99
4.5	Conclusion . . . . .	109
	<b>References</b>	<b>111</b>
	<b>Appendix A Chapter 2 appendices</b>	<b>117</b>
	Appendix A1: Ramsey problem and dimensionality reduction . . . . .	117
	Appendix A2: Calibration details . . . . .	120
	<b>Appendix B Chapter 3 appendices</b>	<b>123</b>
	Appendix B1: Definitions and Ramsey problem . . . . .	123
	Appendix B2: Further results . . . . .	126
	Appendix B3: Solution methods . . . . .	126

---

<b>Appendix C   Chapter 4 appendices</b>	<b>139</b>
Appendix C1: Alternative specifications . . . . .	139
Appendix C2: Further results . . . . .	141
Appendix C2: Calibration . . . . .	141
Appendix C3: Numerical appendix . . . . .	144



# List of figures

2.1	90-10 percentile male earnings ratio changes . . . . .	10
2.2	Finding the individually optimal policy . . . . .	18
2.3	Increase in the individually optimal tax rate . . . . .	19
2.4	Effect of shocks on government assets . . . . .	35
2.5	Simulated economy ( $\alpha_1 = \alpha_2 = 0.5$ ) . . . . .	37
2.6	Increase in government assets/GDP over 30 years . . . . .	38
2.7	Increase in the tax rate over 30 years . . . . .	38
2.8	Consumption equivalence relative to the individually optimal policy . . . . .	41
3.1	Individually optimal tax rate for high and low productivity agent . . . . .	51
3.2	Individually optimal tax rate (with transfers) . . . . .	52
3.3	Increase in individually optimal tax rate (with transfers) . . . . .	53
3.4	Effect of shocks on the tax policy in a static environment . . . . .	60
3.5	Optimal policy over 30 years without shocks ( $\alpha_1 = \alpha_2 = 0.5$ ) . . . . .	61
3.6	Increase in the tax rate over 30 years . . . . .	62
3.7	Increase in transfers over 30 years . . . . .	63
3.8	Consumption equivalence relative to the individually optimal policy . . . . .	64
3.9	Optimal increase in the tax rate as a function of the difference in growth rates . . . . .	66
3.10	Optimal increase in transfers as a function of the difference in growth rates . . . . .	67
4.1	Autarky vs. partial union (policy) . . . . .	102
4.2	Autarky vs. partial union (allocation) . . . . .	103
4.3	Net external debt/GDP of country 1 quantiles in 0.05 steps (partial union) . . . . .	104
4.4	Partial union vs. full union (policy) . . . . .	106
4.5	Partial union vs. full union (allocation) . . . . .	107
B.1	Optimal policy over 30 years ( $\alpha_1 = 0.001, \alpha_2 = 0.999$ ) . . . . .	127
B.2	Optimal policy over 30 years without shocks ( $\alpha_1 = 0.99, \alpha_2 = 0.01$ ) . . . . .	128

C.1	Optimal portfolio with normalisation $b_1^2 = 0$ . . . . .	142
C.2	Optimal portfolio with normalisation $b_2^1 = 0$ . . . . .	143



# List of tables

4.1	Differences between the models . . . . .	92
4.2	Calibration of shocks . . . . .	98
4.3	Fluctuations in consumption and tax rate (autarky or purely aggregate shocks)	100
4.4	Fluctuations in consumption and tax rate in the partial union . . . . .	105
4.5	Fluctuations in assets in the partial union . . . . .	105
4.6	Welfare gains relative to autarky (SWF difference) . . . . .	108
4.7	Welfare gains relative to autarky (consumption equivalence) . . . . .	108
A.1	Summary of calibration . . . . .	122
C.1	Summary of calibration . . . . .	144
C.2	Utility function parameters . . . . .	144



# Chapter 1

## Introduction

This thesis is of the three article format. In this introductory chapter, I briefly outline the relation between the articles, their content, and contribution. Each article is presented in a separate chapter, which also contains a more detailed introduction and discussion of the related literature and empirical facts.

All three articles contribute to the vast literature on optimal fiscal policy following the seminal contributions of Lucas and Stokey (1983) (LS) and Aiyagari et al. (2002) (AMSS). In this literature, a government finances exogenous stochastic expenditures, taxes are distortionary, and markets are either complete (LS), or incomplete (AMSS). A social planner (Ramsey planner) chooses the government policy to maximise a social welfare function.

The articles are concerned with a quantitative analysis of optimal fiscal policy when agents are heterogeneous and markets are incomplete. The first two articles focus on heterogeneity between different types of workers within a country. The third article extends the model of AMSS into an open economy setting and studies the optimal policy in the presence of heterogeneity between households and governments in different countries.

The first two articles study the optimal policy when labour productivity of different groups of people (agents) in the economy follows different trends in an extension of the model of AMSS into a setting with heterogeneous agents. In a typical application these groups are low-skilled and high-skilled workers. This analysis is important mainly for two reasons. Heterogeneous productivity trends are empirically relevant, and their effect on the

optimal policy is quantitatively significant. In the several decades since the 1980s, there has been a significant increase in earnings inequality in the vast majority of OECD countries. In labour economics this is often viewed largely as a consequence of long-run trends such as technological change, globalisation, and decreasing power of unions (Autor (2014)), which affected the high-skilled and the low-skilled workers systematically differently.<sup>1</sup> In a neoclassical model, the increase in earnings inequality over the last several decades is therefore more suitably modelled as caused by heterogeneous productivity trends, rather than by idiosyncratic productivity shocks alone. I show in the following two chapters that the optimal policy is significantly affected by heterogeneity in labour productivity trends of magnitude consistent with the observed increase in earnings inequality in recent decades.

In order to understand intuitively why heterogeneous productivity trends affect the optimal policy, imagine that the labour earnings of the high-skilled grow at a higher rate than the earnings of the low-skilled. For a given level of tax revenue, postponing the tax collection redistributes resources from the high-skilled to the low-skilled, because in the future the earnings of the high-skilled constitute a greater fraction of the tax base.

In the first two articles, labour productivity trends are viewed as capturing all relevant factors which affect systematically differently the labour earnings of different groups of people. Throughout this thesis, I often use the terms *labour earnings* and *labour productivity* interchangeably. Productivities are exogenous in the models, and earnings depend endogeneously on labour supply as well as productivity. In the calibrated models in the first two articles the changes in labour supply (over time and in response to shocks) are small and the endogenous earnings are very strongly correlated with the exogenous productivities.

In the first article (*Heterogeneous Labour Productivity Trends and Optimal Fiscal Policy*), I study the optimal policy without transfers, in a model closely related to Shin (2006). Heterogeneity in growth rates of productivity is particularly important for the optimal policy in this setting, because it creates a conflict between agents over the timing of tax collection.

---

<sup>1</sup>Framing the discussion about the increase in earnings inequality in recent decades by exogeneously dividing workers into two types according to their skill level is a common approach in labour economics, and it succeeds in explaining the most salient trends in earnings distribution (Acemoglu and Autor (2011)). I discuss this in more detail in Chapter 2.

For a given path of government expenditures postponing taxation results in an increase in the government debt. I discuss how the model can contribute to explaining the significant increase in the government debt in many advanced economies in recent decades. I numerically analyse the optimal policy with two types of agents: the low-skilled, and the high-skilled. In particular, I show that labour productivity trends calibrated to match the 1980-2010 US data introduce a significant trend component into the time path of the tax rate and the government debt, and make the solution very sensitive to the preferences of the social planner. For the majority of choices of Pareto weights in the social planner's objective function, the optimal policy entails a significant increase in the government debt. Furthermore, the change in the tax rate and the government debt due to productivity trends in the horizon of several decades is usually greater than the expected change due to government expenditure shocks (precautionary savings of the government). In contrast to the optimal policy, the welfare of the agents is not strongly affected by the preferences of the Ramsey planner. I also discuss the political economics implications of the model.

In the second article (*Heterogeneous Labour Productivity Trends and Optimal Fiscal Policy with Transfers*), I examine the implications of introducing transfers into the model, following Bhandari et al. (2013). Analysing the model with the addition of transfers shows which conclusions of the first article are robust to this change of the set of policy instruments. Furthermore, understanding the quantitative importance of productivity trends in both settings is important in its own right, because it increases the relevance of the analysis for policy-making considerations. The introduction of transfers implies that the total spending of the government excluding interest payments is no longer exogenous. There is a conflict over both the level and the timing of tax collection, and heterogeneity in initial productivities affects the optimal policy more significantly. I analyse the optimal policy for the same calibration as in the first article, targeting the 1980-2010 evolution of earnings distribution in the US. It is still the case that the change in the optimal tax rate over a period of several decades due to productivity trends is usually greater than the expected change due to shocks, but it is smaller than in the model without transfers for most choices of Pareto weights. As Bhandari et al. (2013) show, the level of the government debt is indeterminate in the model

with transfers, and a decreasing or increasing path of taxation is therefore not associated with a change in the government debt. Both the optimal policy and the welfare of the agents when transfers are available strongly depend on the preferences of the social planner.

A major contribution of the first two chapters is to create a framework which can be used to obtain policy recommendations for various predictions of labour productivity evolution. In contrast, a conventional model without productivity trends implicitly assumes that the productivity of all types of workers in the economy grows at the same rate. I discuss the policy-making implications in Chapter 3. I show that even a modest expected heterogeneity in growth rates yields significantly different policy recommendations than a model without heterogeneous productivity trends.

The third article (*Open economy optimal fiscal policy*) is the outcome of joint work with Elisa Faraglia and Rigas Oikonomou.

Many papers study optimal fiscal policy in models with stochastic government expenditures and distortionary taxes, and bonds of multiple maturities.<sup>2</sup> Governments can exploit the negative correlation between bond prices and deficits to achieve some degree of fiscal insurance. In times of adverse expenditure shocks, the market value of the outstanding debt tends to decrease, and issuing long debt can help to ease the adverse expenditure shock. In a closed economy setting, Angeletos (2002) and Buera and Nicolini (2004) show that governments can effectively complete the markets with risk-free bonds of multiple maturities by issuing long term debt.

In an open economy with frictionless capital markets, a no-arbitrage constraint equates the return on bonds of different countries. Equiza-Goni et al. (2016) show in a three-period model that this constraint prevents the government from completing the markets by using fiscal insurance, unless shocks are perfectly positively correlated. The same authors (Equiza-Goni et al. (2017)) present empirical evidence which supports the view that in an open economy, fiscal insurance is possible only when shocks are aggregate. In light of these findings, it becomes particularly interesting to study optimal fiscal policy when markets are incomplete, in the setting of an open economy with frictionless capital markets.

<sup>2</sup>See for example Faraglia et al. (2014a) for an overview of this *debt management* literature.

We examine the implications of extending the closed economy model of optimal fiscal policy with incomplete markets of Aiyagari et al. (2002) (AMSS) into an open economy setting with two countries. Each country is a representative household production economy in the style of AMSS, and its government finances exogenous expenditures, and issues one-period risk-less bonds. We study three different institutional settings: autarky (two closed economies), *partial union* (financial markets are integrated), and *full union* (international transfers are allowed). We study the optimal policy in a calibrated model with ex-ante identical countries and equal Pareto weights, and compare the policy and welfare in the three settings. Since the autarky setting corresponds to the standard model of AMSS, this allows us to analyse the quantitative implications of departing from the closed economy setting. The results in the partial union setting share many features with the closed economy model of AMSS. In particular, the allocation depends on the history of shock realisations, and the consumption random walk result of AMSS still holds. However, there are many important differences. For example, when the correlation of expenditure shocks is perfectly negative, consumption is much smoother over time than in the closed economy, but labour supply fluctuates more. The model also highlights the problem of asymmetry in the spending needs of the governments if they borrow from each other, which results in a greater level of tax distortion in the future.

We solve the optimal fiscal policy problem in a multi-country setting using a global solution method (value function iteration), as opposed to the commonly used approach of solving an LQ approximation of the Ramsey problem around a deterministic steady state. A global solution method is essential in this setting due to the sizeable and persistent effect of shocks on the assets, and the net external debt of the countries in particular.

Solving the optimal policy models with heterogeneous agents and incomplete markets in this thesis is a challenging task for a variety of reasons, such as the presence of forward-looking constraints, non-convexity of feasible sets, a relatively large number of state variables in recursive formulations of the problems, and the presence of trends (in the first two articles). My approach is based on parallel value function iteration and utilises high-performance

computing.<sup>3</sup> Even though there are substantial differences between the three articles, the basic building blocks of the solution algorithms are the same. I discuss the solution methods in detail in the second article (Chapter 3), which also contains an extensive numerical appendix relevant to all three articles.

A significant contribution of the thesis lies in writing the codes for solving the models. The codes are available online in a public GitHub repository:

- [github.com/kvasnicka/phdthesis](https://github.com/kvasnicka/phdthesis)

The rest of this thesis is organised as follows. Each of the following three chapters is dedicated to one of the three articles. There are separate appendices which are placed together at the end of the thesis: Appendix A for Chapter 2, Appendix B for Chapter 3, and Appendix C for Chapter 4.

---

<sup>3</sup>I obtained the numerical results in this thesis using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service. The programs are written in Fortran (2008 standard) and I use Intel Fortran compiler (16.0) which produces highly optimised machine code. The performance gain from this approach relative to using interpreted languages such as Matlab depends on the problem, the quality of the code, hardware, and software. In my experience, the performance gain is very substantial. I had initially attempted to solve the model in Chapter 2 in Matlab. After I rewrote the code in Fortran, the execution time per value function iteration decreased by a factor of around 50-60 when executed on a single processor. Parallelisation of the program resulted in a decrease of time per iteration by a further factor of around 240 (depending on parameters) when executed on 256 processors on the Darwin Supercomputer concurrently.



## **Chapter 2**

# **Heterogeneous labour productivity trends and optimal fiscal policy**

### **2.1 Introduction**

How a government should optimally use distortionary taxes and debt in order to finance its expenditures over time has been a long-standing question in macroeconomics, and a large optimal fiscal policy literature emerged following the seminal articles of Barro (1979) and Lucas and Stokey (1983). Much of the more recent literature following Aiyagari et al. (2002) (AMSS) focused on how a government should optimally respond to expenditure shocks when markets are incomplete, with particular emphasis on the implications of shocks for the optimal policy in the very long run. AMSS showed that in a representative agent economy with quasilinear preferences the optimal policy involves the government accumulating a large stock of precautionary savings, with the tax rate on labour income eventually converging to zero. The convergence is very slow, typically taking hundreds of years.

Across a wide range of specifications of models and their calibrations, the optimal government debt is either significantly negative (as in AMSS) or relatively close to zero, and the convergence to this long-run outcome is slow. Bhandari et al. (2016) studied extensions to AMSS in a representative agent framework such as allowing the single asset in the economy to be risky, introducing more general preferences, and ruling out transfers. In calibrations of

their model targeting US data, the long-run savings of the government are still positive as in the quasilinear example of AMSS, but much closer to zero, and taxes are far from zero in the long run. The rate of convergence to this level of debt is again very slow, typically even slower than in AMSS (the half-life of convergence ranges from 250 to 1200 years depending on details of calibration). Shin (2006) showed that if heterogeneous households are exposed to idiosyncratic productivity shocks, the precautionary savings motive of the government and of households are in conflict, and the savings of the government are positive but close to zero. The general theme of the literature is that in the horizon of years or several decades, the expected change in tax policy and government debt due to expenditure shocks is quantitatively small, and perhaps slightly negative in the case of government debt.

In the last several decades there was a significant increase in debt/GDP ratios in the vast majority of developed economies, which is difficult to reconcile with the literature. Between 1980 and 2010, the average central government debt to GDP ratio in OECD countries increased from 35% to 69%. The median ratio increased from 28% to 47%. The magnitude and timing varied substantially, but an increase in government debt occurred in almost all OECD countries.<sup>1</sup> If a country faces high realisations of government expenditures for many years, it can accumulate a relatively large stock of debt according to a standard model (AMSS), and the increase in debt could possibly be reconciled with the literature as a result of chance. However, this explanation is not empirically plausible. The increase in debt of the magnitude observed in the data (34% of GDP on average over a period of 30 years) is unlikely to be optimal in the case of a single country,<sup>2</sup> and therefore extremely unlikely to be optimal on average in many countries, which do not face perfectly correlated expenditure shocks. Why the vast majority of governments acted seemingly at odds with the normative recommendations of the optimal fiscal policy literature remains a puzzle.

<sup>1</sup>These statistics and comments are based on using data for countries whose data were available for the whole period of 1980-2010 in OECD.Stat online database on 11/03/2013.

<sup>2</sup>In Section 2.4 I solve a standard model of optimal policy (without heterogeneity in productivity trends) for a calibration targeting 1980-2010 US data. The sample probability of debt/GDP ratio increasing by 30 percentage points or more due to government expenditure shocks is 4 percent, with the median *decrease* being almost 20 percentage points.

In this article, I offer a partial explanation of this puzzle. The existing optimal policy models neglect trends in the distribution of labour earnings (increase in income inequality). If we take these trends into account, the optimal policy can involve a quantitatively substantial increase in government debt, and the observed increase in debt may have been much closer to the optimal policy than previously thought. This approach is motivated by the following theoretical and empirical considerations.

Niepelt (2004) shows that if labour productivities of different people in the economy grow at different rates, the optimal policy may entail an increase in debt over time, depending on the preferences of the social planner. The intuition is that the people whose productivity grows relatively slowly, prefer postponing tax collection, because in the future their income will constitute a smaller share of the tax base, and a greater share of any collected tax revenue will be paid by others. Postponing taxation entails an increase in government debt.

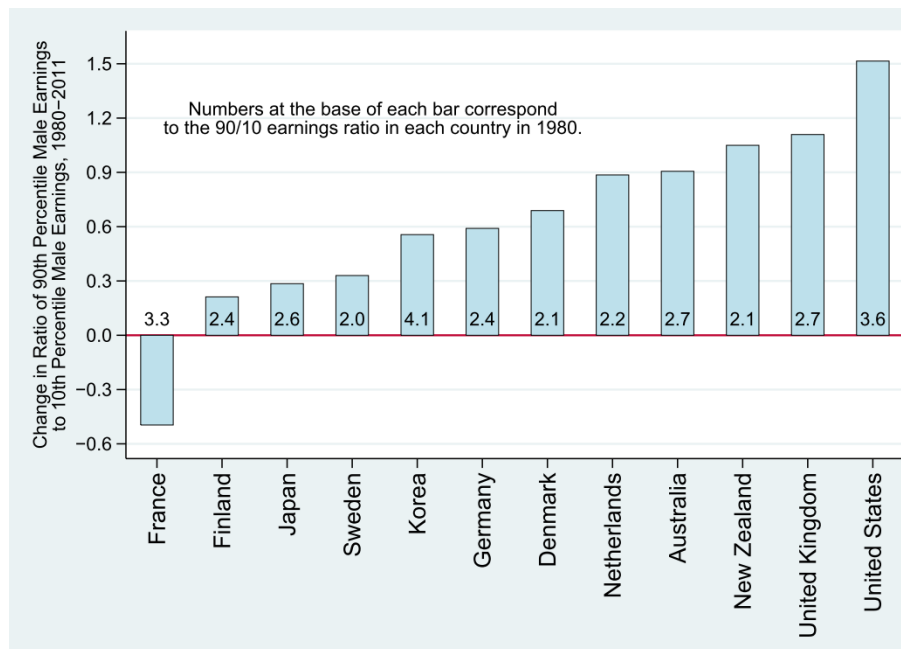
On the empirical side, between 1980s and 2010s there was a significant increase in labour earnings inequality in the vast majority of OECD countries as evidenced by many metrics, such as the 90/10 earnings ratio increase (Autor (2014)). Figure 2.1 shows the increase in 90/10 earnings ratios in various OECD countries. This period of increasing earnings inequality approximately coincided with the period of increasing government debt. Importantly, this increase in earnings inequality was to a large extent a consequence of long-run trends such as technological change, globalisation, and decreasing power of unions, and these trends affected different types of workers systematically differently.<sup>3</sup> It is therefore more appropriate to model the increase in inequality as caused by heterogeneous productivity trends, rather than by idiosyncratic shocks alone.

In this chapter, I show that the productivity trends observed in many advanced economies in recent decades were sufficiently pronounced to provide an explanation for the observed increases in government debt as an optimal response to increasing earnings inequality. More specifically, I introduce agents ex-ante heterogeneous in labour productivity *trends* into a

---

<sup>3</sup>Hornstein et al. (2005), Lemieux (2008), Acemoglu and Autor (2011), and Autor (2014) describe the increase in inequality in the US and other advanced economies, and discuss possible explanations. Technological change and the related rise in college wage premium are a prominent driving force behind the increase in inequality in their analysis. Atkinson (2008) discusses the empirical evidence for OECD countries.

Fig. 2.1 90-10 percentile male earnings ratio changes



[Source: Autor (2014), reprinted with permission from AAAS.]

version of Aiyagari et al. (2002) style production economy, and numerically solve the model calibrated to match US data between 1980 and 2010. I focus on the case of two types of agents; one with low initial productivity and growth rate of productivity (the 'low-skilled' type), and one with high initial productivity and its growth rate (the 'high-skilled' type). This calibration results in an increase in earnings inequality over time. The exogenous division of workers into two types according to their skill level is frequently used in labour economics to model changes in earnings distribution in recent decades, and it explains the most salient trends in data well.<sup>4</sup>

In order to study the effect of productivity trends on policy, in isolation from the effect of shocks, I solve a deterministic calibration of the model. The optimal policy strongly depends on Pareto weights. If the social planner cares only for the welfare of the low-skilled agents,

<sup>4</sup>Acemoglu and Autor (2011) call this approach the *canonical model* and survey its empirical applications. Atkinson (2008) calls it the *textbook model*. The high-skilled types are commonly interpreted as college graduates, and the low-skilled as high school graduates. In recent decades job polarisation has been an important phenomenon in many advanced economies. This means that the share of workers in low skill, low wage jobs has been increasing, and this has also been the case for the share of workers in high skill, high wage jobs (Acemoglu and Autor (2011)). This makes the two-type calibration increasingly empirically relevant.

the tax rate increases by 40 percentage points over a 30-year period, and the government accumulates debt (340% of GDP). If the social planner cares only for the high-skilled, the tax rate decreases by about 6 percentage points over the same period, and the government accumulates assets (37% of GDP). For intermediate Pareto weights, we obtain intermediate results. In particular, an increase in government debt to GDP ratio of 41 percentage points over 30 years is optimal for Pareto weight 0.8 on the high-skilled agent (Pareto weight 0.2 on the low-skilled). We can thus explain the observed increase of debt to GDP in the US of 36 percentage points as optimal for a Pareto weight on the high-skilled close to 0.8.

These results are relevant not only for the US but also for the majority of other developed economies, in which earnings inequality also increased in the last several decades. The increase in inequality was usually less pronounced than in the US. Other things being equal, this means that the optimal change in debt is smaller, and an increase in debt as great or greater than the increase observed in data is optimal for a narrower range of Pareto weights.

The model allows us to compare the relative importance of productivity trends and aggregate shocks in determination of the optimal policy. I solve a stochastic calibration with government shocks targeting stylised facts about US business cycle, and labour productivities fixed at the initial level. The government uses debt to reduce fluctuations in the tax rate in response to shocks. However, in the horizon of several decades the change in policy due to labour productivity trends usually (for most choices of Pareto weights) quantitatively dominates the *expected* change in policy due to expenditure shocks which has been the focus of much of the literature since AMSS. If we focus on the long-run implications of shocks, we may draw wrong conclusions about the nature of the optimal policy. For example, a slight upward trend in government assets may be optimal from the perspective of the precautionary savings motive of the government which was described by AMSS (and from the perspective of other models which do not contain productivity trends). However, this neglects the fact that the low-skilled agents prefer a policy which is associated with a substantial *downward* trend in government assets. Since the low-skilled agents consume less than the high-skilled, this concern should be important to a social planner who cares about equitable outcomes.

This article is most closely related to two papers which analyse optimal fiscal policy without transfers in an economy with agents heterogeneous in labour productivity: Shin (2006) with incomplete markets (one-period bond only), and Bassetto (2014) with complete markets. In this article, the main difference is the addition of heterogeneity in productivity trends. Niepelt (2004) uses a four-period (one per decade) deterministic model in which agents' productivities grow in time at different rates, in order to analyse the implications of reunification of Germany for the optimal policy. As in my model, the government can increase the welfare of some agents at the expense of others by changing the timing of tax collection. In contrast, I analyse a stochastic economy with infinite horizon, more realistic preferences, and a more general process for trend productivity. All of these innovations make my model more empirically relevant.

Several papers follow the seminal contribution of Aiyagari and McGrattan (1998) and quantitatively characterise the optimal level of government debt in a stationary distribution in models with heterogeneous agents and borrowing constraints. The government debt serves to effectively loosen these constraints by providing additional liquidity. The optimal level of government debt tends to be positive in this literature, which could potentially explain the recent widespread increase in government debt as socially optimal. More recently Röhrs and Winter (2017) solve a model for an arguably more realistic calibration of heterogeneity, and conclude that the optimal level of government debt is actually negative. Furthermore, they show that transitions to reach the optimal level of government debt should be very gradual. This literature does not seem to provide a good candidate explanation for the substantial and relatively fast increase in government debt in the majority of developed economies in recent decades.

This article and all the papers mentioned so far assume that the social planner can commit to a policy in the first period for all future periods. For example, Debortoli et al. (2017) study the implications of relaxing this assumption, in the context of a model with multiple maturities of government debt. Analysing the implications of the government's inability to commit for the optimal policy in the context of heterogeneous productivity trends remains a topic for future research.

Broadly, there are two approaches to dealing with the puzzle of increasing government debt, which is the fact that observed policies are at odds with the *seemingly* optimal policy (obtained using a conventional model). The first approach (normative) is to examine whether the seemingly optimal policies are actually optimal once additional time-specific features of the environment are taken into account, and this is the approach I adopt in chapter.<sup>5</sup>

The second approach (positive) is to take as given that governments do not follow optimal policies, and attempt to explain why that is the case. Alesina and Perotti (1994) review classical models in political economics, which aim to explain budget deficits. The class of models which are closest to mine are models of intergenerational redistribution (Tabellini (1991), and Grossman and Helpman (2001)). These models include additional elements relevant for determining the outcome of a political process, but are much more stylised in other aspects and are not well suited for a quantitative analysis. One problem when using these models to explain the increase in debt is explaining why it happened in so many countries, and why in recent decades in particular, even though the demographics and political institutions vary significantly even between developed economies.<sup>6</sup> Even in spite of this problem, these models provide important insights, and the strong dependence of the solution of the model in this article on Pareto weights suggests that political economics considerations are very important in explaining the observed policies. I discuss this at more length in Section 2.5, which is devoted to political economics implications of the model.

The remainder of this chapter is organised as follows. I begin the analysis of the impact of productivity trends on the optimal policy in a two-period simplified model with exogenous tax distortion in Section 2.2. Section 2.3 sets out the full model of optimal policy with heterogeneous productivity trends. I use the model to study the implications of labour

---

<sup>5</sup>I focus on the role of heterogeneous productivity trends. In the same spirit, applying the basic idea of tax-smoothing of Barro (1979), according to which the tax rate should be approximately constant and the government should run up debt in times of temporarily high expenditures or low income, the puzzle of increasing debt could be explained if the majority of governments between 1980s and 2010s expected a downward trend in government expenditures, or an upward trend in aggregate productivity. This explanation is problematic in the light of demographic trends in advanced economies, which are widely expected to adversely affect labour force participation and government expenditures (Bloom et al. (2010)). The expected productivity growth would have to be substantial for the increase in government debt observed in data to be optimal.

<sup>6</sup>In contrast, the increase in inequality coincides with the increase in debt in both space (the majority of advanced economies) and time (particularly pronounced since 1980s).

productivity trends observed in the US in 1980-2010 for the optimal policy in Section 2.4. Political economics implications are discussed in Section 2.5. Section 2.6 concludes. Appendix A1 contains a derivation of the social planner's problem and a recursive formulation of the problem with dimensionality reduced by one. Appendix A2 discusses the calibration strategy in more detail.

## 2.2 Two-period model

It is useful to explore the impact of heterogeneous growth on the optimal policy in a simpler two-period environment before proceeding to the full infinite horizon general equilibrium model. The model presented in this section simplifies the full model by assuming a deterministic environment, small open economy, fixed labour supply, and an exogenous function capturing the distortionary effect of taxation.<sup>7</sup> These simplifications allow us to easily demonstrate the qualitative implications of heterogeneous labour productivity trends.

The government of a small open economy finances its exogenous expenditures by issuing debt and linearly taxing labour income. There are two periods indexed by  $t = 0, 1$ , and  $I$  agents indexed by  $i$ , with mass  $\pi_i > 0$ ,  $\sum_i \pi_i = 1$ . Agent  $i$ 's utility is  $U_i = u(c_{i,0}) + \beta u(c_{i,1})$ , where  $c_{i,t}$  is  $i$ 's consumption in period  $t$ ,  $u' > 0, u'' < 0$ , and  $\beta \in (0, 1)$ .

The production technology is a linear function of labour input. Agents differ in their initial productivity of labour and in its growth rate. Agent  $i$ 's initial labour productivity is  $\theta_{i,0}$ , which is the amount of output that one unit of labour supplied by agent  $i$  in period 0 produces, and growth rate of productivity is  $\xi_i$ , so  $i$ 's productivity in the second period is  $\theta_{i,1} = (1 + \xi_i)\theta_{i,0}$ . All agents supply one unit of labour in each period so the aggregate product in period  $t$  is  $Y_t = \sum_i \pi_i \theta_{i,t}$ .

---

<sup>7</sup>The model is essentially a two-period version of the representative agent model of Barro (1979), with agents heterogeneous in the initial productivity and growth rate of productivity. Another difference is that Barro directly assumes that the objective of the government is to minimise the tax distortion, whereas I assume that the government maximises a weighted sum of the agents' utilities. In the representative agent setting these assumptions are equivalent. With heterogeneous agents they are equivalent (in the sense of equivalence of the optimal policy) only in the special case in which the productivity of all agents grows at the same rate.



Both the agents and the government can save and borrow without limit at a given gross interest rate  $R > 0$ . The government has to finance its exogenous expenditures of present value  $G > 0$ , and it raises the revenue to do so by taxing labour income in the two periods at rates  $\tau_0, \tau_1$ . Following Barro (1979), taxation is associated with distortion (or collection costs), and this is modelled by an exogenous distortion function  $D(\tau_t, Y_t)$ . Distortion in period  $t$  depends only on the tax rate  $\tau_t$  in the period, and on the aggregate production (tax base)  $Y_t$ . We further assume  $\frac{\partial D}{\partial \tau} > 0, \frac{\partial^2 D}{\partial \tau^2} > 0$ .<sup>8</sup> The government's budget constraint is

$$G \leq \tau_0 Y_0 - D(\tau_0, Y_0) + \frac{1}{R} (\tau_1 Y_1 - D(\tau_1, Y_1)), \quad (2.1)$$

and agent  $i$ 's budget constraint is

$$c_{i,0} + \frac{1}{R} c_{i,1} \leq (1 - \tau_0) \theta_{i,0} + \frac{1}{R} (1 - \tau_1) \theta_{i,1}, \quad (2.2)$$

for  $i = 1, \dots, I$ . The assumption that all agents and the government have zero initial and final asset holdings does not affect the conclusions of this section as long as the government's asset holdings are sufficiently small that it needs to collect taxes in order to satisfy its budget constraint.

I assume that the distribution of individual productivities and growth rates is such that the aggregate product  $Y_t$  is bounded in both periods and there is a *feasible tax policy*, that is a tax policy for which the government's budget constraint and the budget constraints of all agents are satisfied.<sup>9</sup>

Agent  $i$ 's utility maximisation problem is to choose consumption  $c_{i,0}, c_{i,1}$  in the two period, which maximises her utility  $U_i$  subject to her budget constraint (2.2), taking the tax policy  $(\tau_0, \tau_1)$  as given. A competitive equilibrium given a tax policy  $(\tau_0, \tau_1)$  is a consumption allocation  $\{c_{i,t}\}_{i,t}$  which solves the utility maximisation problem of all agents, and the budget constraints of all agents and the government are satisfied.

<sup>8</sup>These assumptions mean that for a given aggregate income (tax base)  $Y$ , the distortion increases in the tax rate, and at an increasing pace. The sign of  $\frac{\partial^2 D}{\partial Y^2}$  is left unspecified as it is not important for the results discussed here. In Barro (1979) this was a constant positive number as the distortion was assumed to be CRS in the tax base and the overall tax collection (rather than the tax rate).

<sup>9</sup>In general, this might not be the case if  $G$  is sufficiently great or the distortion  $D$  is sufficiently strong.

The *socially optimal policy* (Ramsey plan) is a tax policy  $(\tau_0, \tau_1)$  for which a competitive equilibrium allocation results in the maximum possible social welfare  $\sum_i \alpha_i U_i$ , where  $\alpha_i \geq 0, i = 1, \dots, I$  are Pareto weights. Any socially optimal policy in this model involves the same present value of tax collection (net of distortion) equal to the exogenous level of government expenditures.<sup>10</sup> In other words, there is no conflict between agents over the *level* of tax collection.

### 2.2.1 The individually optimal policy

Due to the heterogeneity in growth, there is a conflict over the *timing* of tax collection. To see this it is useful to analyse *individually optimal policies*, which are the socially optimal policies with a positive Pareto weight on one agent and zero on every other agent. The individually optimal policy of agent  $i$  is a policy  $(\tau_{0,i}^*, \tau_{1,i}^*)$  that maximises the present value of  $i$ 's after-tax labour income. This is a consequence of the assumption that agents face a single present-value budget constraint, so their utility is an increasing function of the present value of their after-tax income.

Figure 2.2 graphically represents the problem of finding an individually optimal policy, for the example of a quadratic<sup>11</sup> distortion function  $D(\tau, Y) = aY\tau^2, a > 0$ . The *efficient policies* curve depicts all combinations of tax rates in the two periods for which the government's budget constraint is satisfied with equality and distortion is minimised (we are not 'past the peak of Laffer curve' which can happen due to convex distortion). All socially optimal policies, and individually optimal policies in particular, must lie on this curve.<sup>12</sup> Its slope

<sup>10</sup>In other words the government's budget constraint is binding in a socially optimal equilibrium. If it were not the case then the government could decrease the tax rate in some period which would increase the welfare of all agents and still satisfy the government's budget constraint.

<sup>11</sup>The quadratic distortion was chosen for illustrative purposes only and it is not meant to be empirically plausible. All results discussed here generalise to other distortion functions  $D$  as long as these satisfy the properties  $\frac{\partial D}{\partial \tau} > 0, \frac{\partial^2 D}{\partial \tau^2} > 0$ .

<sup>12</sup>If a socially optimal policy did not lie on this curve, it would mean that we can decrease the tax rate in one or both periods while still satisfying the government's budget constraint. Such a change would yield an increase in utility of every agent (as long as the agent has a strictly positive income in both periods), which contradicts the assumption that we started at a socially optimal policy. Even agents with zero income in some period would not strictly benefit from deviating from this curve, so we can restrict our attention to policies on the efficient policies curve.

determines the change in the tax rate in the second period for a given change in the tax rate in the first period so that the present value of tax revenue net of distortion remains the same.

An indifference curve of an agent depicts combinations of tax rates which yield a constant level of after-tax income. Their slope equals  $-R/(1 + \xi_i)$ , which crucially depends only on the growth rate of the agent's productivity and not on her *level* of productivity per se. This is a consequence of the linearity of the tax schedule.<sup>13</sup> The agent's individually optimal policy is at a point of tangency between the indifference curve closest to the origin (which is the direction of increasing utility), and the efficient policies curve. I plot the indifference curves and the corresponding optimal policies for two agents, one with a high productivity growth rate ( $\xi_i = 1.0$ ) and one with a low rate ( $\xi_i = 0.5$ ). The agent whose productivity grows at a relatively high rate prefers a relatively high tax rate in the first period and a low tax rate in the second period. Distortion serves as a moderating force which makes deviations from a flat tax profile increasingly costly. In its absence the efficient policy curves would be linear and agents would choose corner solutions, i.e. they would want all of the tax collection to be done in the period in which their income is relatively low.

In the general case of a strictly increasing and convex distortion function, it can be shown<sup>14</sup> that if there is an interior solution (tax rates are between 0 and 1 in both periods), agent  $i$ 's optimal policy is determined by the equation

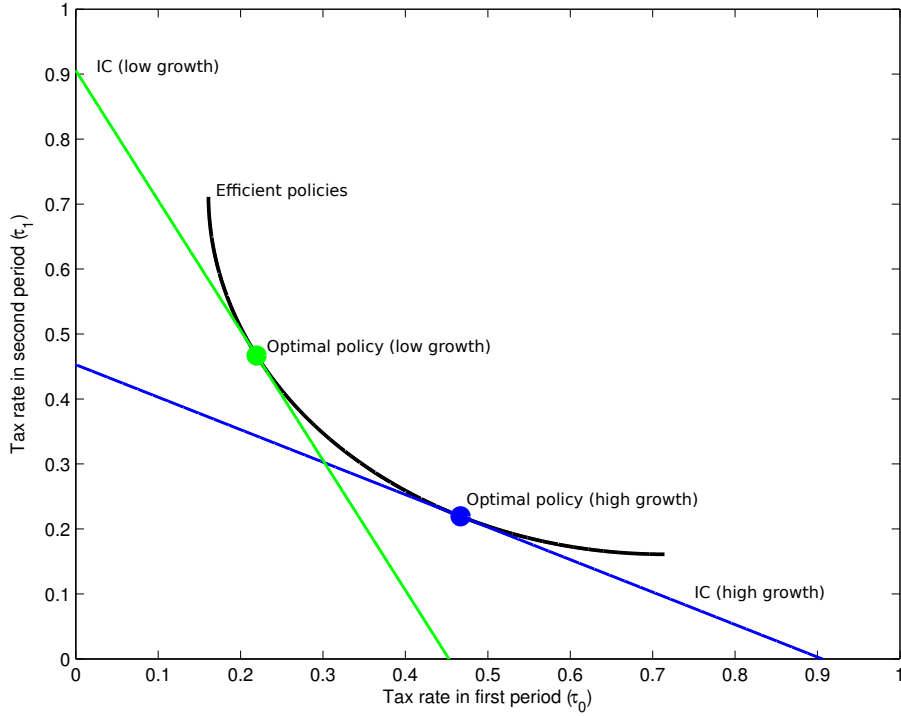
$$1 + \xi_i = \frac{\theta_{i,1}}{\theta_{i,0}} = \frac{Y_1 - \frac{\partial D(\tau_1, Y_1)}{\partial \tau_1}}{Y_0 - \frac{\partial D(\tau_0, Y_0)}{\partial \tau_0}} \equiv \frac{\text{MTR}_1}{\text{MTR}_0} \quad (2.3)$$

and the government's budget constraint (2.1). Equation (2.3) captures the distributional conflict over the timing of taxation which arises due to heterogeneity in growth rates. It states that the ratio of agent  $i$ 's productivities (gross growth rate) equals the ratio of the marginal

<sup>13</sup>An increase in the tax rate in a period results in a change in agent's after-tax income which is proportional to her pre-tax income (productivity) in that period. If the tax rate in period 0 increases by  $\Delta\tau_0$ , the present value of agent  $i$ 's after tax income increases by  $-\Delta\tau_0\theta_{i,0}$ . To keep her present value of income constant (and thus keep her on the same indifference curve) we need to increase the tax rate in the second period by  $\Delta\tau_1 = -R\frac{\Delta\tau_0\theta_{i,0}}{\theta_{i,1}}$ , which depends only on the ratio of productivities, not their level.

<sup>14</sup>Using the first-order conditions of the problem of maximising agent  $i$ 's after tax income, subject to the government's budget constraint.

Fig. 2.2 Finding the individually optimal policy



tax revenue in periods 1 and 0. The ratio of productivities reflects the relative cost of paying taxes in the two periods from the perspective of the agent, and the ratio of the marginal tax revenues determines the change in the tax rate in the two periods which are necessary to keep the tax revenue constant. If this does not hold, we can find an alternative policy which increases the utility of the agent.<sup>15</sup> Going back to the graphical interpretation, this equation states that the slope of the efficient policy curve in Figure 2.2 equals the slope of the indifference curve of agent  $i$ .

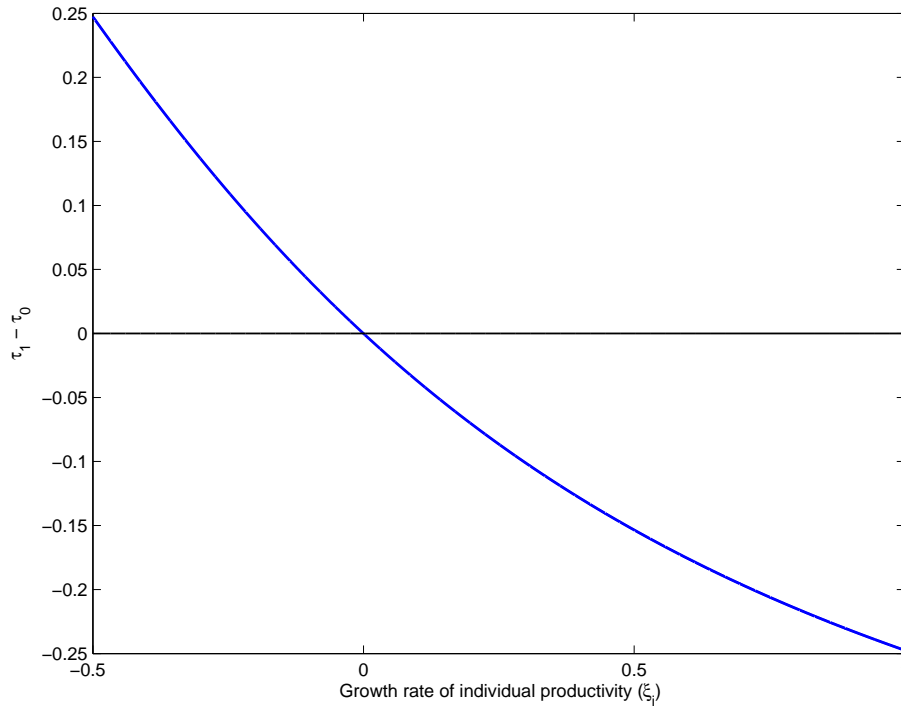
Holding the aggregate product  $(Y_0, Y_1)$  constant, changing the *level* of productivity of agent  $i$  does not affect  $i$ 's individually optimal policy, since  $i$ 's productivity enters the equation only as the ratio of productivities in the two periods. It follows that if two agents have the same growth rate of productivity, their individually optimal policies are the same.

<sup>15</sup>Suppose that, starting at the individually optimal policy, the social planner increases the tax rate in the first period by  $\Delta\tau_0$ . To keep the present-value of the government expenditures unchanged, the tax rate increase in the second period must be  $\Delta\tau_1 = -R \frac{MTR_0}{MTR_1} \Delta\tau_0$ . The impact of this change on the present value of agent  $i$ 's income is  $\Delta PV_i = \Delta\tau_0 \left( \theta_{i,0} - \theta_{i,1} \frac{MTR_0}{MTR_1} \right)$ . If equation (2.3) does not hold, then we can find an alternative policy which increases the present value of  $i$ 's income by setting  $\Delta\tau_0 > 0$  or  $\Delta\tau_0 < 0$ .

Suppose on the other hand that two agents  $i$  and  $j$  have a different productivity growth rate, and that  $i$ 's productivity grows faster, i.e.  $\xi_i > \xi_j$ . It can be shown<sup>16</sup> that in this case  $\tau_{1,i}^* < \tau_{1,j}^*$ , and  $\tau_{0,i}^* > \tau_{0,j}^*$ . The agents whose productivity grows relatively fast prefer a combination of a low tax rate in the future ( $t = 1$ ) and a high tax rate in the present ( $t = 0$ ), which is associated with an increase in government debt between the periods.

Figure 2.3 shows the increase in the tax rate between the two periods corresponding to the individually optimal policy of agent  $i$  as a function of  $i$ 's growth rate of productivity for the quadratic distortion example.<sup>17</sup>

Fig. 2.3 Increase in the individually optimal tax rate



<sup>16</sup>From equation (2.3) the optimal policies satisfy

$$\frac{Y_1 - \frac{\partial D(\tau_{1,i}^*, Y_1)}{\partial \tau_1}}{Y_0 - \frac{\partial D(\tau_{0,i}^*, Y_0)}{\partial \tau_0}} > \frac{Y_1 - \frac{\partial D(\tau_{1,j}^*, Y_1)}{\partial \tau_1}}{Y_0 - \frac{\partial D(\tau_{0,j}^*, Y_0)}{\partial \tau_0}}.$$

Because  $D$  is increasing and strictly convex in  $\tau$ , this inequality (and the government's budget constraint satisfied with equality) implies that  $\tau_{1,i}^* < \tau_{1,j}^*$ , and  $\tau_{0,i}^* > \tau_{0,j}^*$ .

<sup>17</sup>For more general distortion functions which satisfy the assumptions on continuity and derivatives stated earlier, the increase in the tax rate is still a continuous and decreasing function of the growth rate of agent  $i$ 's productivity.

I now address the implications of these results for the *socially* optimal policies in which the Pareto weight is positive on more than one agent.

**Homogeneous growth** If there is no heterogeneity in growth ( $\xi_i = \xi$  for all  $i$ ), then the individually optimal policy is the same for all agents. It follows that the socially optimal policy does not depend on the Pareto weights or on the distribution of initial productivities (holding the aggregate product constant). It minimises the present value of distortion, which is the 'tax-smoothing' result of Barro (1979). If  $D$  is further CRS in  $Y$  as in Barro's model, the optimal policy entails the same tax rate in both periods.

**Heterogeneous growth** If there are at least two agents with different growth rate of productivity ( $\xi_i \neq \xi_j$ ), their individually optimal policy is different. The social planner can increase the welfare of some agents at the expense of others (and increase distortion in the process) by deviating from a flat tax profile. The socially optimal policy depends on the distribution of productivity levels, growth rates, and Pareto weights.

The conflict over the timing of tax collection in the two-period model arises because of the heterogeneity in the growth rates. The role of the heterogeneity in levels lies merely in affecting how the conflict is resolved.

### 2.2.2 Limitations of the two-period model

The main goal of this article is to quantitatively analyse the implications of heterogeneous productivity trends which have been observed in many economies in recent decades for the optimal policy, and to assess whether these can explain the observed increases in government debt. The model of this section is far too stylised for this purpose. The problem lies not only in the crude approximation of time by two periods, but, equally importantly, in the exogenous distortion function. Distortion in the model is of critical importance. If it is strong, deviations from a flat tax profile are costly, and the optimal policy will not vary much with Pareto weights, because the gains from redistribution by changing the timing of taxation are to a large extent eliminated due to the high distortionary cost of taxation. Assumptions

about the functional form and parametrisation of distortion would drive the results and would be extremely difficult to justify. To overcome these issues, I set up a standard infinite horizon DSGE model of optimal policy following Aiyagari et al. (2002), in which distortion follows from first principles as endogenous response of agents' labour supply to the tax policy. Another significant advantage of this approach is that it makes the results of this article more directly comparable to the existing literature, as the addition of heterogeneity in trends is the only major departure.

Despite the additional complexity, much of the intuition regarding the individually optimal policies, and by extension, socially optimal policies, carries over from the two-period model. In particular, the intuition about the role of heterogeneity in growth rates (creating a conflict) and heterogeneity in levels (resolving the conflict) is still relevant, and the cases of extreme choices of Pareto weights (one on one agent, zero on the other) resemble the individually optimal policies in the two-period model.

## 2.3 The full model

### 2.3.1 The environment

I follow the branch of literature on optimal fiscal policy with incomplete markets started by the representative agent model of Aiyagari et al. (2002). The most closely related model to mine is the heterogeneous agents model of Shin (2006).<sup>18</sup> There are  $I$  types of infinitely lived agents indexed by  $i = 1, \dots, I$ . The mass of agent of type  $i$  is  $\pi_i$ , with  $\sum_{i=1}^I \pi_i = 1$ . Agent  $i$ 's objective function (at  $t = 0$ ) is defined over stochastic sequences of consumption  $\{c_{i,t}\}_{t=0}^{\infty}$  and labour  $\{l_{i,t}\}_{t=0}^{\infty}$ :

$$\mathbb{E}_0 \left[ \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right]. \quad (2.4)$$

---

<sup>18</sup>The main differences are that in my model there are heterogeneous productivity trends, and there are no idiosyncratic productivity shocks.

$U^i$  is the instantaneous utility function,  $\beta \in (0, 1)$  is the discount factor, and  $\mathbb{E}_t$  denotes expectation conditional on information available in period  $t$ . By assumption,  $l_{i,t} \in [0, \bar{l}_i]$ , where  $\bar{l}_i > 0$  is given. Derivatives of the utility functions with respect to consumption and labour are denoted as  $U_c^i$  and  $U_l^i$ ,  $U_i$ 's are twice continuously differentiable,  $U_c^i > 0, U_l^i < 0, U_{cc}^i < 0, U_{ll}^i < 0$ .

The consumption good is perishable and it is produced by CRS technology using labour input only. One unit of labour supplied by type  $i$  agent in period  $t$  produces  $\theta_{i,t}$  units of output, so the aggregate resource (feasibility) constraint is

$$\sum_{i=1}^I \pi_i c_{i,t} + g_t \leq \sum_{i=1}^I \pi_i \theta_{i,t} l_{i,t} \quad (2.5)$$

for  $t = 0, 1, \dots$ , where  $g_t$  is government expenditure at  $t$ .

The adjusted government expenditure  $\{\bar{g}_t\}_{t=0}^\infty$  is

$$\bar{g}_t = k \sum_{i=1}^I \pi_i \theta_{i,t} \bar{l}_i \quad (2.6)$$

for  $t = 0, 1, \dots$  and some given  $k > 0$ . The interpretation is that the adjusted government expenditure at time  $t$  is a constant share of the maximum possible output at time  $t$  (the output if all agents spent their whole time endowment working). If the agents choose to supply approximately constant amount of labour in all periods in equilibrium, the equilibrium ratio of the adjusted government expenditure to aggregate product will be approximately constant over time.<sup>19</sup> The actual government expenditure at  $t$  is

$$g_t = s_t \bar{g}_t \quad (2.7)$$

for  $t = 0, 1, \dots$ , where  $s_t$  follows a finite-state Markov process, with  $s_t > 0$  for all  $t$ , and transition probabilities are denoted as  $Pr(s_{t+1}|s_t)$ .  $s^t = (s_0, \dots, s_t)$  denotes a history of shocks up to period  $t$ . A multiplicative shock specification is more appropriate in the presence of

<sup>19</sup>This assumption simplifies the calibration, and the ratio of  $g/Y$  approximately constant is empirically plausible for the periods which I study. The model can easily be generalised so that the adjusted government expenditure is an arbitrary function of time as the productivities of the agents.



trends than additive shocks, because the magnitude of additive shocks as share of the economy would vary significantly over time.

The heterogeneous productivity trends are modelled as exogenous functions,

$$\theta_{i,t} = \theta_i(t), i = 1, \dots, I. \quad (2.8)$$

A special case useful for ease of calibration and interpretation of the results is the case of constant growth rates of individual productivities  $\xi_{\theta,i} > -1$ , so that

$$\theta_{i,t+1} = (1 + \xi_{\theta,i})\theta_{i,t}, \quad (2.9)$$

for  $t = 0, 1, \dots$  and all  $i$ , with initial productivities  $\theta_{i,0} > 0$  given. The functions  $\theta_i$  can be easily generalised to depend on  $s_t$  so we could analyse the impact of idiosyncratic shocks within this model.

Functions  $\theta_i(t)$  become constant in  $t$  after  $\mathcal{T}$  periods:  $\theta_i(t) = \theta_i(\mathcal{T})$  for  $t \geq \mathcal{T}$ .  $\mathcal{T}$  can be large but finite. This assumption is very convenient for solving the model, and it is not limiting in the study of the impact of heterogeneous productivity trends in recent decades on the optimal policy.<sup>20</sup>

Markets are incomplete. The only asset in the economy is one-period risk-free bond issued by the government or the agents.  $b_{i,t}$  denotes holdings of this bond by agent  $i$  bought at time  $t$ , maturing at time  $t + 1$ . Let  $R_t$  denote a gross interest rate from time  $t$  to  $t + 1$ . Labour income of every agent in period  $t$  is taxed at a proportional tax rate  $\tau_t$ . Agent  $i$ 's sequence of budget constraints is

$$c_{i,t} + b_{i,t} \leq (1 - \tau_t)\theta_{i,t}l_{i,t} + R_{t-1}b_{i,t-1}, \quad (2.10)$$

for  $t = 0, 1, \dots$ . Without loss of generality, assume that  $R_{-1} = 1$ .

---

<sup>20</sup>I discuss why the assumption is convenient in section 2.4. In the empirical applications of the model, the growth is not fast enough that the returns in late periods  $t \geq \mathcal{T}$  significantly affect policy in early periods. I used  $\mathcal{T} = 30$  in the computations of the numerical examples in this article. Solving the model for cases in which growth continues further does not change the policy in the first decades significantly.

Let  $B_t$  denote the government's assets purchased in period  $t$ , maturing in period  $t + 1$  ( $-B_t$  is government's debt). The government's budget constraint is

$$g_t + B_t \leq \tau_t \sum_{i=1}^I \pi_i \theta_{i,t} l_{i,t} + R_{t-1} B_{t-1}, \quad (2.11)$$

for  $t = 0, 1, \dots$

Agents and the government start with initial holdings of assets  $\{b_{i,-1}\}_{i=1}^I$ ,  $B_{-1}$ , and the market-clearing condition at every period  $t \geq -1$  is<sup>21</sup>

$$\sum_{i=1}^I \pi_i b_{i,t} + B_t = 0. \quad (2.12)$$

As the majority of the optimal policy literature, I focus on interior solutions with respect to any borrowing constraints the agents might face. This simplifies the analysis considerably. Following Bhandari et al. (2013), I assume that the assets of the agents are bounded from below. This economises on the notation relative to the alternative common approach of assuming explicit borrowing constraints, which are loose enough that they never bind in equilibrium.

The social planner's objective function is

$$\mathbb{E}_0 \left[ \sum_{i=1}^I \pi_i \alpha_i \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right], \quad (2.13)$$

where  $\alpha_i > 0$  is Pareto-weight on agent  $i$  and  $\sum_i \alpha_i = 1$ .

### 2.3.2 Competitive equilibrium

I now define the competitive equilibrium and related concepts.

---

<sup>21</sup>It is possible for the agents to borrow from one another even without the involvement of the government ( $B_t = 0$ ). This is not the case in an environment with a representative household, in which the household can only borrow from the government.

**Definition** (*allocation, asset profile, price system, tax policy*)

An *allocation* is a sequence  $\{c_{i,t}, l_{i,t}\}_{i,t}$ . An *asset profile* is a sequence  $\{\{b_{i,t}\}_i, B_t\}_t$ . A *price system* is a sequence  $\{R_t\}_t$  of interest rates on risk-free 1-period bonds. A *tax policy* is a sequence of tax rates on labour income  $\{\tau_t\}_t$ .

**Definition** (*competitive equilibrium*)

For given initial assets held by the agents  $\{b_{i,-1}\}_i$  and by the government ( $B_{-1}$ ), and a given tax policy  $\{\tau_t\}_t$ , a *competitive equilibrium* is an allocation  $\{c_{i,t}, l_{i,t}\}_{i,t}$ , an asset profile  $\{\{b_{i,t}\}_i, B_t\}_t$ , and a price system  $\{R_t\}_t$ , such that:

1. The allocation  $\{c_{i,t}, l_{i,t}\}_{i,t}$  and assets held by agents  $\{b_{i,t}\}_{i,t}$ , maximise all agents' utility (2.4) subject to agent-specific budget constraints (2.10).
2. Resource constraint (2.5) is satisfied.
3. The government's budget constraint (2.11) is satisfied (this is redundant by Walras' law).
4. Market-clearing in the asset market (2.12) is satisfied.
5. The asset profile is bounded.

**Definition** (*socially optimal competitive equilibrium - Ramsey plan*)

Given the initial distribution of assets  $(\{b_{i,-1}\}_i, B_{-1})$ , a *socially optimal competitive equilibrium* (Ramsey plan) is a tax policy  $\{\tau_t^*\}_t$ , an allocation  $\{c_{i,t}^*, l_{i,t}^*\}_{i,t}$ , an asset profile  $\{\{b_{i,t}^*\}_i, B_t^*\}_t$ , and a price system  $\{R_t^*\}_t$  such that:

1. Given the initial asset holdings  $(\{b_{i,-1}\}_i, B_{-1})$  and the tax policy  $\{\tau_t^*\}_t$ , the other elements of the socially optimal competitive equilibrium  $\{c_{i,t}^*, l_{i,t}^*\}_{i,t}$ ,  $\{\{b_{i,t}^*\}_i, B_t^*\}_t$ ,  $\{R_t^*\}_t$  constitute a competitive equilibrium.
2. Given the initial asset holdings  $(\{b_{i,-1}\}_i, B_{-1})$ , there is no other tax policy  $\{\tau_t\}_t \neq \{\tau_t^*\}_t$ , for which there is a competitive equilibrium which yields higher value of social welfare (2.13).

The socially optimal equilibrium is a solution of a standard Ramsey problem in which the objective function of social planner is (2.13). I refer to the elements of a *socially optimal competitive equilibrium* as a socially optimal tax policy, allocation, price system, and asset profile.

### 2.3.3 The Ramsey problem

In order to solve the Ramsey problem I follow the standard primal approach. This means that I use the first-order conditions of the agents' utility maximisation problems to derive implementability constraints which do not depend on the tax rate and the interest rate, but on the allocation and the asset profile only. Once we solve for the optimal allocation, the optimal tax policy and interest rate can be recovered using the first-order conditions of any of the agents' problems. Appendix A1 contains a detailed derivation of the Ramsey problem:

$$\max_{\{c_{i,t}(s^t), l_{i,t}(s^t), b_{i,t}(s^t)\}_{t=0, \forall i, s^t}, \{R_t(s^t)\}_{t=0, \forall s^t}} \mathbb{E}_0 \left[ \sum_{i=1}^I \pi_i \alpha_i \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right]$$

subject to

$$c_{i,t}(s^t) + b_{i,t}(s^t) = - \frac{U_l^i(c_{i,t}(s^t), l_{i,t}(s^t))}{U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))} l_{i,t}(s^t) + R_{t-1}(s^{t-1}) b_{i,t-1}(s^{t-1}), \forall i, t, s^t \quad (2.14)$$

$$\frac{U_l^i(c_{i,t}(s^t), l_{i,t}(s^t))}{\theta_{i,t} U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))} = \frac{U_l^j(c_{j,t}(s^t), l_{j,t}(s^t))}{\theta_{j,t} U_c^j(c_{j,t}(s^t), l_{j,t}(s^t))}, \forall i, j, t, s^t \quad (2.15)$$

$$\sum_{i=1}^I \pi_i c_{i,t}(s^t) + g_t(s^t) = \sum_{i=1}^I \pi_i \theta_{i,t} l_{i,t}(s^t), \forall t, s^t \quad (2.16)$$

$$R_t(s^t) = \frac{U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))}{\beta \mathbb{E}_t[U_c^i(c_{i,t+1}, l_{i,t+1})]}, \forall i, t, s^t \quad (2.17)$$

$$(2.18)$$

The agents' budget and resource constraints bind in a competitive equilibrium because the utility functions are strictly increasing in consumption and strictly decreasing in labour. The

government's budget constraint becomes redundant by Walras' law. Constraint (2.17) follows from the focus on equilibria in which no agent is borrowing-constrained.

### 2.3.4 Recursive formulation of the Ramsey problem

In order to solve the model I use a recursive formulation of the Ramsey problem. The optimal policy is in general not time-consistent. I follow the approach of Kydland and Prescott (1980) who were the first to propose a recursive formulation to deal with this problem, by including a pseudo-state variable (marginal utilities), and splitting the problem into time  $t = 0$  and time  $t \geq 1$  problems. In the context of optimal policy models this approach was used by Shin (2006), Farhi (2010), or Golosov and Sargent (2012).

The value function for periods  $t \geq 1$  is an ex-ante value function. In other words, it captures the expected return at the start of the period before the current-period shock is realised, and we find a state-contingent plan for the current period. The notation is standard in the sense that for any variable  $x$ ,  $x'$  denotes the next-period value and  $x_-$  the last-period value. The state vector is  $(\mathbf{b}, \mathbf{u}, t, s_-)$ :

- $\mathbf{b} \equiv (b_1, \dots, b_I)$  are assets purchased by the agents in the last period, maturing in the current period.
- $\mathbf{u} \equiv (u_1, \dots, u_I)$  are marginal utilities of consumption of all agents in the last period.
- $t$  is 'time' governing the evolution of productivities  $\theta_i(t)$ .
- $s_-$  is the last-period shock realisation.

The choice variables are  $[c_i(s), l_i(s), b'_i(s)]$  for every possible realisation of shock  $s$  and for all agents  $i$ . The summation  $\sum_s$  means summation over all possible shock realisations.

Notation  $\mathbf{b}'(s) \equiv (b'_1(s), \dots, b'_I(s))$  for the vector of current-period asset purchases and  $\mathbf{U}_c(s) \equiv (U_c^1[c_1(s), l_1(s)], \dots, U_c^I[c_I(s), l_I(s)])$  for the vector of current-period marginal utilities is used. I also write  $U^i(s) \equiv U^i(c_i(s), l_i(s))$  and use the same notation for marginal utilities.  $\mathbb{E}_{s_-}[\cdot]$  denotes expectation with respect to distribution of  $s$  conditional on  $s_-$ .

The Bellman equation for periods  $t \geq 1$  is

$$V(\mathbf{b}, \mathbf{u}, t, s_-) = \max_{[c_i(s), l_i(s), b'_i(s)]_{i,s}} \sum_s Pr(s|s_-) \left[ \sum_{i=1}^I \pi_i \alpha_i U^i(c_i(s), l_i(s)) + \beta V(\mathbf{b}'(s), \mathbf{U}_c(s), t', s) \right]$$

subject to

$$c_i(s) + b'_i(s) + \frac{U_l^i(c_i(s), l_i(s))}{U_c^i(c_i(s), l_i(s))} l_i(s) = \frac{1}{\beta} \frac{u_i}{\mathbb{E}_{s_-}[U_c^i(c_i, l_i)]} b_i, \forall i, s \quad (2.19)$$

$$\frac{\mathbb{E}_{s_-}[U_c^i(c_i, l_i)]}{u_i} = \frac{\mathbb{E}_{s_-}[U_c^j(c_j, l_j)]}{u_j}, \forall i, j \quad (2.20)$$

$$\frac{U_l^i(c_i(s), l_i(s))}{\theta_i U_c^i(c_i(s), l_i(s))} = \frac{U_l^j(c_j(s), l_j(s))}{\theta_j U_c^j(c_j(s), l_j(s))}, \forall i, j, s \quad (2.21)$$

$$\sum_{i=1}^I \pi_i c_i(s) + g = \sum_{i=1}^I \pi_i \theta_i l_i(s), \forall s \quad (2.22)$$

$$\theta_i = \theta_i(t), \forall i \quad (2.23)$$

$$t' = \min\{t + 1, \mathcal{T}\} \quad (2.24)$$

$$g = \bar{g}s, \forall s \quad (2.25)$$

$$\bar{g} = k \sum_{i=1}^I \pi_i \theta_i \bar{l}_i \quad (2.26)$$

Constraints (2.19) are the implementability constraints following from the agents' utility maximisation first-order conditions and their budget constraints. Constraints (2.20) and (2.21) follow from the agent's first-order conditions when used to eliminate the rate of return and tax rate, and the assumption that all agents face the same tax rate and rate of return. (2.22) is the aggregate resource constraint. Equation (2.23) yields all agents' individual productivities from the 'time' state variable  $t$ . Equation (2.24) is the law of motion for 'time', which is incremented by one unless it already reached the upper bound  $\mathcal{T}$  in which case it is unchanged. Constraints (2.25) give the actual government expenditure as a function of  $\bar{g}$  and shock realisation  $s$ , and constraint (2.26) gives the adjusted government expenditures as a function of productivities. Appendix A1 contains an alternative formulation which reduces

the number of state variables by one, which is very useful in numerically solving the model accurately.

The period 0 problem is different as the last-period marginal utilities are not defined:

$$(c_0^*, l_0^*, b_0^*) = \arg \max_{\{c_{i,0}, l_{i,0}, b_{i,0}\}_i} \sum_{i=1}^I \pi_i \alpha_i U^i(c_0, l_0) + \beta V(\mathbf{b}_0, \mathbf{U}_c, t_1, s_0)$$

subject to

$$\begin{aligned} c_{i,0} + b_{i,0} + \frac{U_l^i(c_{i,0}, l_{i,0})}{U_c^i(c_{i,0}, l_{i,0})} l_{i,0} &= b_{i,-1}, \forall i \\ \frac{U_l^i(c_{i,0}, l_{i,0})}{\theta_{i,0} U_c^i(c_{i,0}, l_{i,0})} &= \frac{U_l^j(c_{j,0}, l_{j,0})}{\theta_{j,0} U_c^j(c_{j,0}, l_{j,0})}, \forall i, j \\ \sum_{i=1}^I \pi_i c_{i,0} + g_0 &= \sum_{i=1}^I \pi_i \theta_{i,0} \bar{l}_i \\ t_1 &= \min\{1, \mathcal{T}\} \\ g_0 &= \bar{g}_0 s_0, \forall s \\ \bar{g}_0 &= k \sum_{i=1}^I \pi_i \theta_{i,0} \bar{l}_i \end{aligned}$$

The initial asset holdings  $\{b_{i,-1}\}_{i=1}^I, B_{-1}$  and initial shock  $s_0$  are given and consistent with the period  $t = -1$  market-clearing condition for assets. Notation  $\mathbf{b}_0 = (b_{1,0}, \dots, b_{I,0})$  and  $\mathbf{U}_c = (U_c^1(c_{1,0}, l_{1,0}), \dots, U_c^I(c_{I,0}, l_{I,0}))$  is used. Note that there is no expectation operator before the continuation value. This is the case because the value function for period  $t \geq 1$  captures the value before the current-period shock is realised.

### 2.3.5 Solution methods

In this section, I outline the solution algorithm briefly. A broader discussion of my choice of the solution approach, alternative approaches, and implementation details is contained in Chapter 3 and its numerical appendix.

The computational strategy is to use a recursive formulation of the problem to find an approximation of period  $t \geq 1$  value function. Once we have an approximation of the value

function, solving the first-period problem for any initial state starting from period  $t = 0$  and simulating the consecutive periods is straightforward. I do not use the recursive formulation of subsection 2.3.4 directly, but rather a transformation derived from it, which reduces the number of state variables by one, and which is given in Appendix A1. It utilises the fact that the marginal utilities appear in the original recursive formulation only as the ratio of one another, and as a multiplier of each agent's asset holdings. In the appendix I give the formulation for the case of two agents. The reduction of dimensionality can be generalised to an arbitrary number of agents (see Bhandari et al. (2013)).

The vector of state variables after transformation is  $(a_1, a_2, \rho, t, s_-)$ , where  $a_i = \frac{1}{\bar{p}} u_i b_i$  for  $i = 1, 2$  are marginal-utility adjusted asset holdings of both agents, and  $\rho = \frac{u_1}{u_2}$  is the ratio of their marginal utilities of consumption.

I used discrete value function iteration to obtain the results reported in this chapter. The value function is approximated as a list of values for every point on a grid covering a portion of the state space, and continuation values in between grid points are obtained by multi-dimensional interpolation.

Optimal policy problems with heterogeneous agents and incomplete markets are difficult to solve due to a combination of a relatively large number of state variables (hence the number of grid points is great due to curse of dimensionality), and the complexity of the problem at a given point in state space. What makes the model of this article particularly challenging to solve is the presence of an extra state variable ( $t$ ) compared to Shin (2006), and the fact that the productivity trends induce a significant variation over time in the state variables, so the grids need to cover a larger subset of the state space than would otherwise be necessary. For a given number of grid points, if we make the grid wider, the density of the grid and thus accuracy of the solution is reduced, hence we need a large number of grid points to obtain an accurate solution.

Fortunately, value function iteration is a problem which naturally lends itself to parallelisation. In order to solve the model accurately, I wrote a parallel program in Fortran which is well-scalable on hundreds of processors on a supercomputer.



I ran the computations to obtain the results in the following section on the Darwin cluster of High Performance Computing service of University of Cambridge. I solved the model using various number of grid points and processors. To illustrate the number of grid points and computation time, I give two examples. I solved the model with deterministic calibrations that are presented on the following pages using 200 grid points for asset holdings of every agent, 100 grid points for the ratio of marginal utilities, and 30 grid points for the 'time' state variable, giving a total of 120 million grid points. Running the program on 128 processors in parallel, it took 12 hours to complete 130 iterations in the value function iteration algorithm. I solved a stochastic version of the model using 180 million grid points, taking 24 hours to finish 160 iterations on 256 processors.

The stochastic version of the model is particularly challenging to solve accurately, and I implemented accuracy tests based on Euler equation residuals in the first-order equations in the social planner's problem. For a calibration with productivity fixed at the initial level (which allows us to take a relatively narrow and thus dense grid), the average EE residual in terms of consumption was 0.3 percent. For a stochastic calibration with growth (in which I had to use a less dense grid), the average EE residual in terms of consumption change was 1.2 percent.

At this stage, I address the computational usefulness of the assumption of potentially great but finite number of periods  $\mathcal{T}$  for which the agent's productivities grow. The first advantage is that if the growth stops eventually, we can use a denser grid because there is less variation in state variables over time, and we obtain a more accurate solution. The second advantage is related to the possibility of choosing a recursive formulation with a compact state space. If the 'time' variable  $t$  keeps growing ad infinitum, then for any fixed grid boundary it eventually leaves the grid. In state  $t = \mathcal{T}$  we would then need to use extrapolation to obtain a continuation value, and this would most likely lead to divergence in the value function iteration.

## 2.4 Increasing inequality and optimal fiscal policy

In this section, I use numerical solutions of the model to analyse the implications of the increase in earnings inequality since the 1980s for the optimal policy, with the emphasis on the puzzle of increasing government debt (in the vast majority of industrial countries in recent decades). I calibrate the model on 1980-2010 US data, but the results are also relevant to other economies in which earnings inequality also increased.

I proceed in two steps. Firstly I show that for a calibration which corresponds to standard models (by keeping the labour productivity of every agent fixed at its initial level), the model cannot explain the increase in debt as a response to government expenditure shock.

I then show that once we calibrate the model to match the observed increase in inequality, it can account for the increase in government debt. The optimal policy in the face of increasing inequality strongly depends on the preferences of the social planner. The optimal increase in government debt to GDP ratio is between minus 37 and plus 340 percentage points over 30 years depending on the choice of Pareto weights. For equal Pareto weights on all agents the increase in debt is 156 percent of GDP.

### 2.4.1 Calibration

The calibration targets stylised data on US economy between 1980s and 2010s. I solve the model for a calibration with two agents. Agent 1 is 'high-skilled'. Her productivity is initially high and also grows faster. Agent 2 is 'low-skilled'. Her productivity is initially low and is fixed at the initial level. To summarise,  $\theta_{1,0} > \theta_{2,0}$ ,  $\xi_{\theta,1} > \xi_{\theta,2} = 0$ . This calibration generates a long-run increase in inequality consistent with the data.

The preferences are identical for both agents

$$U^i(c_{i,t}, l_{i,t}) = \log(c_{i,t}) - B \frac{l_{i,t}^{1+\gamma}}{1+\gamma}, \quad (2.27)$$

where  $B, \gamma > 0$ . Assuming logarithmic utility of consumption is common in the optimal policy literature.

The value of Frisch elasticity of labour supply ( $\frac{1}{\gamma}$ ) is 0.5. This value is also used in a quantitative example of Bhandari et al. (2013), even though higher values are often used in macro models (Peterman (2014)). In this model, the value of 0.5 fits data better, as for higher elasticities there would be a quantitatively significant and counterfactual secular trend in hours worked.<sup>22</sup>

The effect of increasing the elasticity of labour supply on the magnitude of the trends in optimal tax rate and government debt is not clear. On one hand it makes deviations from a constant tax rate more costly in terms of distortion (because agents respond more strongly to changes in the tax rate). On the other hand it yields more scope for redistribution by deviating from a flat tax profile, as the earnings of the agents are more concentrated in the periods in which they are more productive. As a robustness check I also solved the model for a Frisch elasticity value of 1. The trends in optimal policy were actually slightly more pronounced than with the baseline value.

The discount factor  $\beta$  is chosen to target 1980-2010 average real interest rate in the US of 5.26 percent.<sup>23</sup> The value of  $B$  is chosen such that the labour supply in a deterministic steady state with productivities fixed at their initial level is 0.7, which is a commonly used value (for example by Faraglia et al. (2008), Faraglia et al. (2010)). The maximum labour supply in a period was set to  $\bar{l} = 3$  following Shin (2006) for stability reasons, so the target is about one quarter of labour endowment.<sup>24</sup>

Agents are divided into two groups of equal mass  $\pi_1 = \pi_2 = 0.5$ . I calibrated the trend productivities functions as constant-growth rate processes. The initial levels and growth rates of productivities are set to match 90-10 earnings ratio in the US in 1980 and 2011 (taken from Autor (2014)), if the labour supply was inelastic. Because of intertemporal substitution of labour, the resulting earnings ratio will slightly deviate from the target. Initial asset holdings

<sup>22</sup>Due to the presence of trends, the agents whose productivity grows at a higher rate, work relatively more in later periods, and vice versa for the others. When Frisch elasticity is 0.5 these trends are relatively flat which is consistent with the data.

<sup>23</sup>Data was downloaded from World Bank World Development Indicators database in May 2015.

<sup>24</sup>The 'stability' lies in the fact that in relatively rare circumstances, at some grid points it is difficult to find a solution since 'travelling' in some directions, the candidate solution rapidly leaves and enters the non-convex feasible sets. Relaxing the constraint on maximum feasible labour helps with this problem to some extent (as there is one less constraint that might be an issue in this), and in simulations these high values of  $l$  are never chosen.

are set to zero for all agents. The productivity of the agent with lower growth rate is fixed at the initial level. Bhandari et al. (2013) used the same calibration strategy for productivity, targeting the 90-10 income ratio with equal mass of agents. The initial assets were set to zero.

The results that the productivity trends are quantitatively significant for policy, and that it is possible to explain the observed increase in debt for some choice of Pareto weights, are not driven by the calibration with two agents only. A calibration with a greater number of agents would involve agents with more extreme values of productivities, and their growth rates. The individually optimal policies of these agents would entail even more pronounced trends in policy (including the change in government debt) than is the case in the two-agent calibration.

To calibrate government expenditures the target tax rate in a steady-state with productivities fixed at their initial levels is  $\tau_{ss} = 0.25$ . This is consistent with 1980-2010 average total federal tax rate of 20.70% in the US<sup>25</sup> adjusted upwards by a population-weighted average of state tax rates.<sup>26</sup>

I solve the model for both deterministic calibrations, in which the government's expenditure multiplicative shock is always  $s_t = 1$ , and stochastic calibrations in which it fluctuates between low ( $s_t = s^L < 1$ ) and high ( $s_t = s^H > 1$ ) levels. In the case of stochastic calibrations, I use the same targets for average output decline in recession, duration of recession, and duration of boom, as Bhandari et al. (2013). These data correspond to stylised facts about US business cycle between late 1980s and 2010.

Appendix A2 contains more detail on the calibration, including a table which summarises all calibration targets and parameter values.

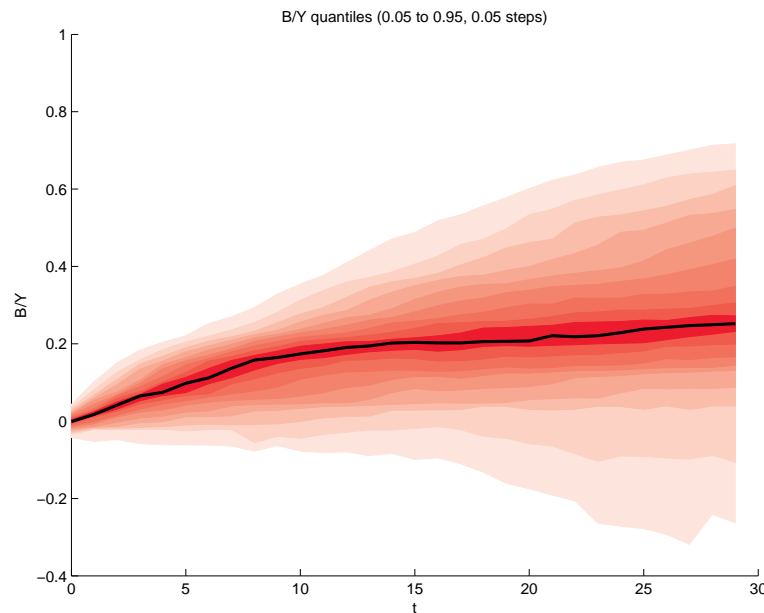
<sup>25</sup>This average was calculated using data from the Congressional Budget Office compiled by the Tax Policy Centre.

<sup>26</sup>The data on state tax rates are average taxes in the fourth quintile of pre-tax income distribution for the year 2007 from Davis (2009), and the population weights are constructed from the 2010 US Census (<http://www.census.gov/2010census>). The fourth quintile was chosen due to availability of data, and in the case of federal tax rate, it is very close to the average. A share of other state taxes than income tax is included to match overall level of taxation better. For shares between 0 and 0.5, this yields target tax rate between 0.232 and 0.266.

### 2.4.2 Results

To begin with, I show that optimal response to government expenditure shocks alone (not taking the heterogeneity in trends into account) cannot explain the observed increase in government debt in most industrial countries since the 1980s. I solve the model for a stochastic calibration in which the productivity of both agents is kept at the initial level, and the Pareto weights are equal on both agents.<sup>27</sup> This calibration corresponds to the existing literature. Figure 2.4 shows quantiles of sample paths of government *assets* to GDP ratio for 1,000 simulations.<sup>28</sup>

Fig. 2.4 Effect of shocks on government assets



The range of possible government assets (debt) after 30 years caused by the expenditure shocks is quite wide. For a particular realisation of shocks, the government debt to GDP can decrease by up to around 70 percentage points, or increase by up to around 25 percentage points. The government engages in tax-smoothing, as it accumulates assets in good times (when expenditures are low) and decreases its asset holdings in bad times. On average, the

<sup>27</sup>When there are no productivity trends the dependence of the optimal policy on Pareto weights is weak (for this calibration), and the results are very similar for a different choice of Pareto weights.

<sup>28</sup>In every period, the figure shows 0.05 to 0.95 quantiles of the government assets to GDP ratio. I permute the initial debt conditions slightly by drawing the agents' initial asset holdings uniformly from interval of length 0.1 centred around the initial conditions.

government debt to GDP ratio *decreases* by around 20 percentage points over the 30 year period, which is a reflection of the precautionary savings motive described by Aiyagari et al. (2002). The result that the effect of shocks on debt can be large for a particular path of shock realisations but is low on average, is consistent with the existing literature as was discussed in the introduction. For the US economy, the sample probability of debt to GDP *increasing* by 30 or more percentage points was around 4 percent. This figure is similar for other countries if their business cycle characteristics were similar to the US business cycle.

Because the probability that a single country optimally increases its government debt to GDP ratio by more than 30 percentage points over a thirty year period is low (4% in the case of the US calibration), the probability that this increase is optimal averaged across many countries which do not face perfectly correlated expenditure shocks is very low. In other words, the significant increase in the average debt/GDP ratio in OECD countries (34 percentage points over 30 years) cannot be explained as an optimal response to expenditure shocks by a standard model with heterogeneity in levels but not in growth rates of productivity.

In contrast, once we take the heterogeneity in productivity trends into account, the model can explain a significant increase in government debt as socially optimal. In order to study the effect of trends on the optimal policy in isolation from the effect of shocks, I solve the model for a deterministic calibration with productivity growth, and a range of Pareto weights  $\alpha_1$  on the high-skilled agent, and  $\alpha_2 = 1 - \alpha_1$  on the low-skilled agent.

Figure 2.5 shows most variables of interest for the simulated economy for Pareto weights equal on both agents.<sup>29</sup> Because of the decreasing marginal utility of consumption, the social planner chooses a policy that redistributes resources from the high-skilled agent to the low-skilled agent. The low-skilled agent whose productivity grows at a lower rate benefits from an increasing tax rate over time, as suggested by the analysis of the individually optimal policies in the two-period model. The tax rate increases by 15.8 percentage points over the 30 year period, and the government's debt to GDP ratio increases by 156 percentage points. The labour supply of the low-skilled workers is decreasing over time, and the labour supply

<sup>29</sup>One of the plots shows transfers, which are not included in this model and are thus always zero. These are introduced into the model in the following chapter.

of the high-skilled is approximately constant.<sup>30</sup> The decrease in overall labour supply as share of time endowment explains the slightly increasing government expenditures as share of GDP, as the adjusted government expenditures are a function of the maximum output (corresponding to a constant labour supply), not the actual output.

Fig. 2.5 Simulated economy ( $\alpha_1 = \alpha_2 = 0.5$ ). Solid lines correspond to the high-skilled agent (1), dashed lines to the low-skilled (2).

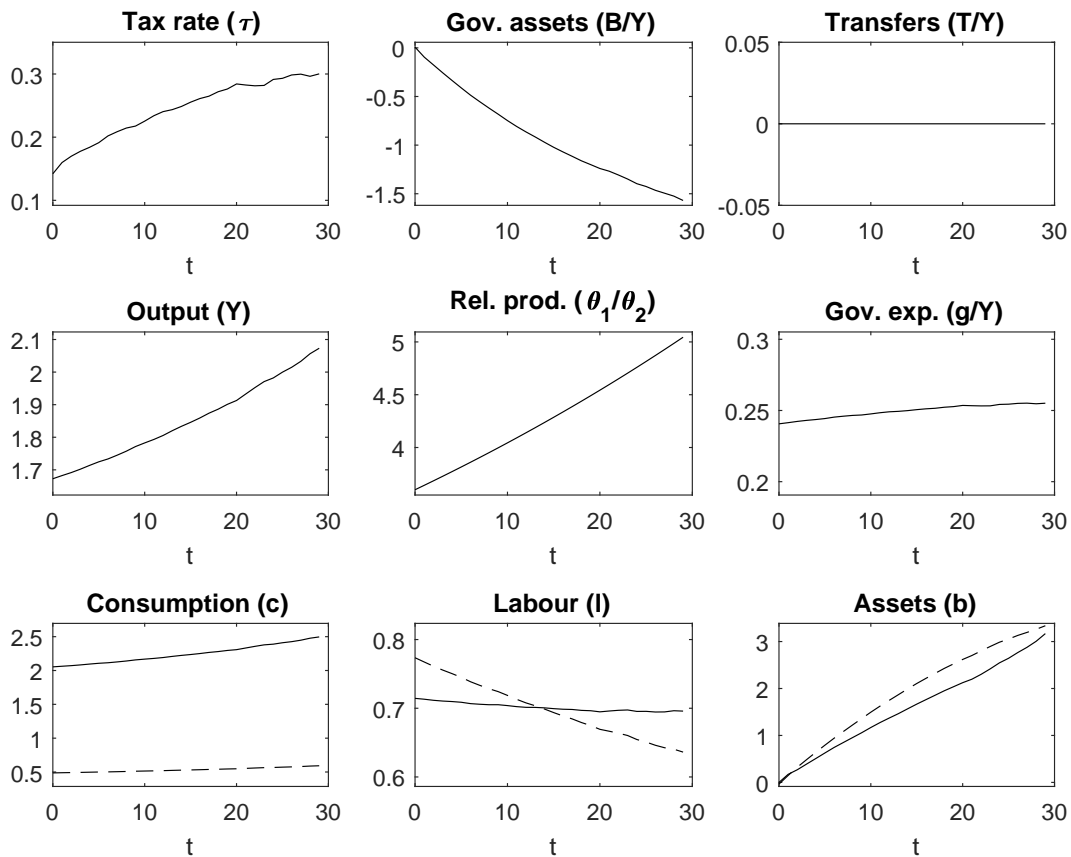


Figure 2.6 shows the increase in government assets (decrease in debt) relative to GDP ( $Y$ ) over a 30 year period as a function of Pareto weight  $\alpha_1$  on the high-skilled agent, with the Pareto weight on the low-skilled agent being  $\alpha_2 = 1 - \alpha_1$ . Positive values mean that the government's assets increased (debt decreased). The changes in debt were accompanied by significant changes in labour taxation over time. Figure 2.7 shows the difference in the tax rate as a function of Pareto weight on the high-skilled agent in the same period.

<sup>30</sup>In a deterministic calibration, we can show analytically that the ratio of labour supply of the high-skilled to the labour supply of the low-skilled is increasing in a socially optimal equilibrium. The levels of labour supply can either increase or decrease.

Fig. 2.6 Increase in government assets/GDP over 30 years

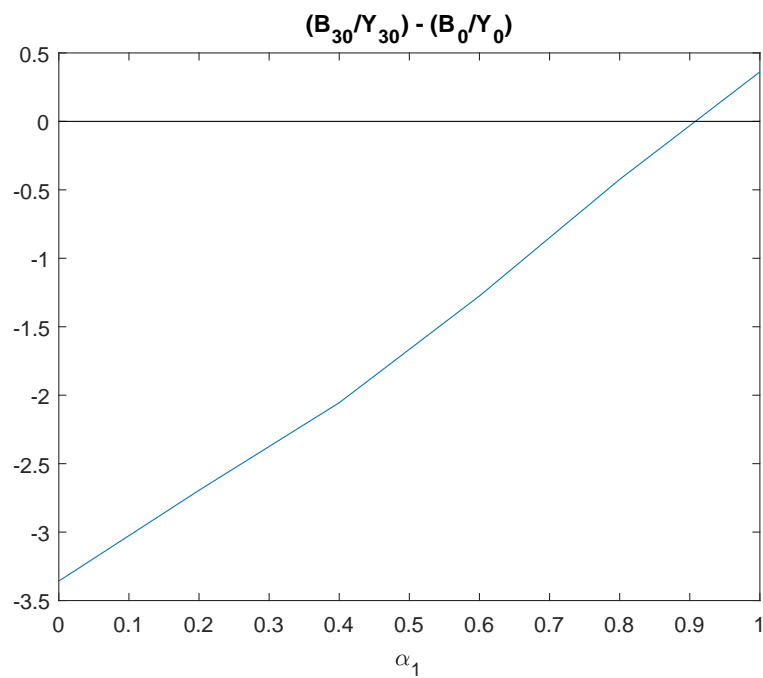
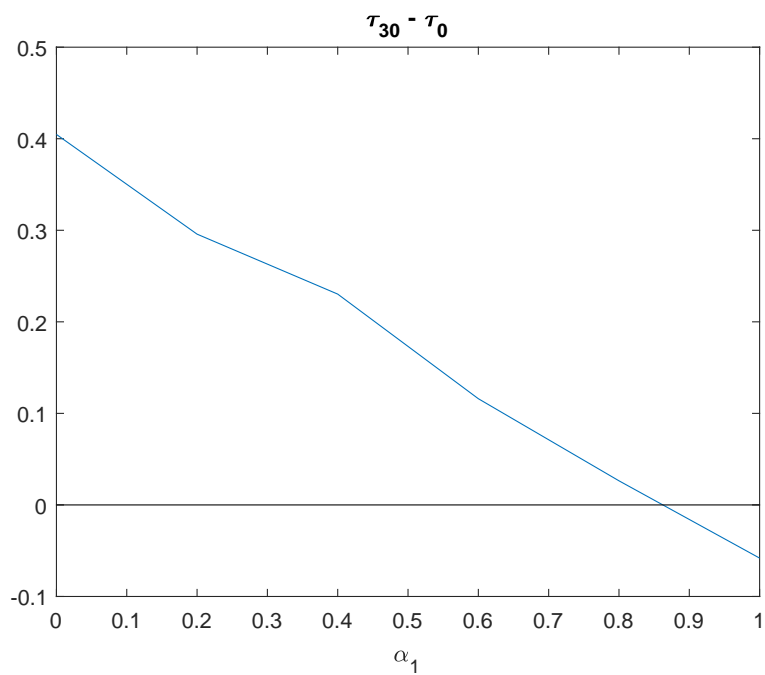


Fig. 2.7 Increase in the tax rate over 30 years





When the Pareto weight on the high-skilled agent is low, it is optimal for the government to increase the tax rate over time, and accumulate debt. As the Pareto weight on the high-skilled agent increases, the optimal increase in the tax rate declines, until it eventually becomes negative, and the government runs up less debt, eventually turning into a net saver. The dependence of the solution on Pareto weights is strong. If the social planner cares only about the low-skilled, the debt to GDP ratio increases by 340 percentage points over 30 years (tax rate increases by 40 percentage points). In the opposite extreme case when the social planner cares only about the high-skilled, the debt to GDP ratio decreases by 40 percentage points (tax rate decreases by 6 percentage points). For the majority of choices of Pareto weights, it is optimal to increase government debt to GDP ratio by the same amount or more than the increase observed in the US in 1980-2010 (of 36 percentage points).

Compared to the relatively small expected change in government debt to GDP ratio in this period due to expenditure shocks (20 percentage points decrease), the effect of productivity trends on government policy is substantially greater for most choices of Pareto weights. Studying optimal responses to shocks at business cycle frequency is important. However, if we focus on the long-run implications of shocks, we may draw conclusions which may be valid in the long run, but are misleading in the horizon of decades. For example, the precautionary savings motive of the government which leads to an (expected) increase in government assets goes against the interests of those whose productivity is growing relatively slowly, and who prefer policies which entail a decrease in government assets.

If the social planner cares only about the low-skilled ( $\alpha_1 = 0$ ), the deviation from a constant tax profile is much greater than if she cares only about the high-skilled. There are two reasons for this asymmetry. Firstly, the income of the low-skilled agent constitutes a low share of the tax base in all periods. The amount of resources which can be extracted from the low-skilled by changing the timing of taxation so as to shift some of the tax burden on them is limited for this reason, and compares less favourably to the distortionary costs of this policy. Secondly, because the income of the high-skilled constitutes a high share of the tax base in all periods, the high-skilled bear the majority of distortionary costs associated with deviating from a flat tax profile.

I also solve the model for calibrations with both shocks and productivity growth in effect at the same time. Preliminary results suggest that the introduction of shocks has little effect on the sample average policy, in the sense that the average of the simulated stochastic solutions is very close to the deterministic solutions; the drift in the expected policy due to the presence of shocks is small relative to the change in policy due to trends. However, the accuracy of the solution in terms of Euler equation residuals is worse, because we need to resort to a less dense grid in this case.<sup>31</sup> Obtaining more accurate solutions in this setting remains a topic for future work.

In the context of the two-period model I concluded that the heterogeneity in growth rates creates a conflict over policy, and the heterogeneity in levels is important in how the conflict is resolved. The same intuition applies to the full model. In a deterministic calibration with constant government expenditure in which the productivity of all agents is fixed at the initial level (there is no heterogeneity in growth), the tax rate is constant over time and the solution does not depend on Pareto weights.<sup>32</sup> The introduction of heterogeneity in growth is what creates a distributional conflict over timing of taxation. In a stochastic calibration, the policy in general depends on Pareto weights, but the conflict is weak in the sense that the policy is very similar for different choices of Pareto weights.<sup>33</sup>

We have seen that the optimal policy strongly depends on the Pareto weights, as the change in the tax rate ranges between minus 6 and plus 40 percentage points. In contrast, the corresponding differences in welfare of the agents are relatively minor. Figure 2.8 shows, for each agent, the welfare cost of moving from the agent's individually optimal equilibrium to a socially optimal equilibrium with Pareto weight  $\alpha_1$  on the high-skilled.<sup>34</sup> It is expressed as a consumption equivalence. A value  $k$  means that the allocation is equivalent in terms of utility

<sup>31</sup>Firstly, there is one more state variable if the productivity is not fixed at the initial level. Secondly, the trends induce more variation in states over time, and the grid needs to be wider in each dimension.

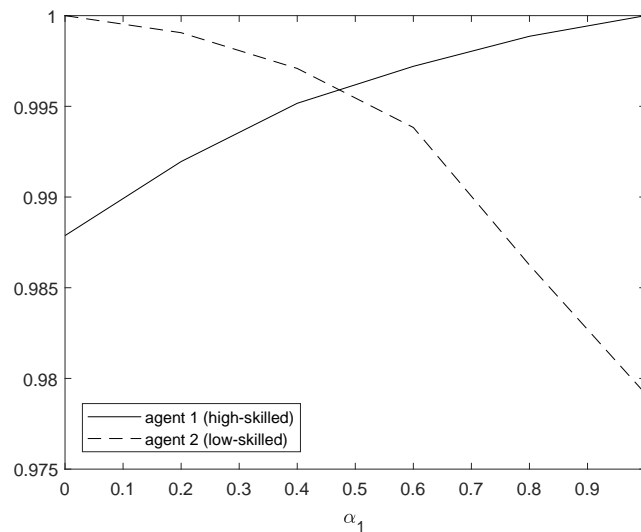
<sup>32</sup>This can be shown analytically using first-order conditions of the planner's problem. The policy is independent of Pareto weights because the level of expenditures is fixed, and every agent wants to choose the lowest tax rate which suffices to pay for the government expenditures.

<sup>33</sup>This statement is true for the calibrations studied in this article, in which the main source of heterogeneity is labour productivity. Bassetto (2014) studies the optimal policy response to expenditure shocks for various choices of Pareto weights if asset holdings are a major source of heterogeneity, in a complete markets setting. In this case, the choice of Pareto weights affects the optimal response to shocks more significantly.

<sup>34</sup>In other words, this reflects the cost of not being a dictator, i.e. not being able to implement one's individually optimal policy, as a function of the Pareto weight of the social planner who chooses the policy.

to the individually optimal allocation, if its consumption element is multiplied by  $k$ . For example, the value of 0.98 for the low-skilled agent when  $\alpha_1 = 1$  means that the low-skilled are indifferent between the allocation which maximises the welfare of the high-skilled, and their own individually optimal allocation in which consumption is reduced by 2 percent. The fact that all these values are close to one can be interpreted as meaning that not much redistribution can be achieved by deviating from a flat tax profile. There are two forces which limit the redistribution potential in this model. Firstly, there is the standard intratemporal distortionary effect of increasing the tax rate (which leads to the 'tax-smoothing' result of Barro (1979)). Secondly, as the social planner attempts to exploit an existing concentration of earnings of different agents in the future or the present, the agents respond by intertemporal substitution of labour away from periods with the increased tax rate.

Fig. 2.8 Consumption equivalence relative to the individually optimal policy



## 2.5 Political economics implications

The social planner's perspective is a useful starting point for explaining observed policies. It demonstrates that in the presence of heterogeneous productivity growth, a wide range of choices of Pareto weights in the social planner's function is consistent with the increase in

debt in advanced economies in recent decades. However, the optimal policy is very sensitive to the choice of Pareto weights, and a wide range of outcomes can be explained as optimal for some choice.

In order to truly understand observed policies in periods characterised by a substantial heterogeneity in productivity growth, we need to adopt a positive perspective. A natural approach is that of political economics, i.e. modelling the preferences of the government explicitly rather than assuming the existence of a benevolent social planner. In the spirit of the new political economics literature, it is useful to first consider which policies are preferred by different individuals, and then model how these preferences are aggregated through a political process.<sup>35</sup>

The model developed in this chapter can be used to quantitatively characterise the individually optimal policies, which are the socially optimal policies with a positive Pareto weight on one agent only. These have a positive interpretation, telling us what is in the best interest of each type of agent. Coupled with assumptions about the political process, this approach can yield testable implications.

If the model with two agents is calibrated such that the mass of agents is not equal, and we assume that the policy for all periods is determined by a majority vote in period  $t = 0$ , the policy pursued by the government will be the individually optimal policy of the type with greater mass.<sup>36</sup>

Consider a setting in which the heterogeneity in productivity trends occurs to a large extent due to trends such as skill-biased technological change, which affect systematically differently the ‘high-skilled’ (college graduates) and the ‘low-skilled’, and inequality is increasing. As was discussed in the introduction, this setting is empirically relevant to developments in the labour market in many advanced economies in recent decades. The model predicts that the high-skilled prefer a decreasing tax rate over time, and the low-skilled

<sup>35</sup>See for example Mueller (2003) for an introduction to the methodology and examples of applications.

<sup>36</sup>If there are more than two types, the analysis becomes significantly more complicated even in the case of voting over the government policy only once in period 0. The policy is infinite-dimensional, and the heterogeneity of agents is three-dimensional (initial productivities, their growth, and initial assets). Piguillem and Schneider (2013) show in a similar setting that if heterogeneity is only one-dimensional (either in initial productivities, or assets) and markets are complete, then the individually optimal fiscal policy of the median agent is a Condorcet winner (it cannot be defeated by any other policy in a pair-wise vote).

prefer an increasing tax rate, and quantifies these policy trends. If the low-skilled constitute the majority of the population, then the model predicts that the government policy coincides with the individually optimal policy of the low-skilled, and entails an increasing tax rate over time, associated with an increase in government debt.

The individually optimal policy of one agent harms the other agent by shifting the tax burden on her. If agents are altruistic, they consider the effect of a policy on the other agents, and the individually optimal policies are closer to each other than suggested by the baseline model without altruism.<sup>37</sup> This will be reflected in their voting over the government policy.

One possible setting in which altruism is a particularly relevant concern is the study of trends which affect different age groups systematically differently. If we calibrate the model to match different productivity trends affecting the young and the old, and we neglect altruism, we may obtain the prediction that these groups prefer a vastly different fiscal policy. Due to standard life-cycle considerations, we can expect the individually optimal policy of the old to involve taxes increasing over time, as their labour earnings can be expected to grow relatively slowly (and then drop after retirement). The young prefer the opposite policy of a decreasing tax rate. However, altruism is arguably strong within families,<sup>38</sup> and can be expected to limit the desire of the young to extract resources from the old or vice versa, by changing the timing of taxation. In this case the individually optimal policies obtained using the model without altruism will overestimate the desire of the agents to deviate from a relatively flat tax profile. To draw quantitatively correct conclusions about the individually optimal policies and the government policy, we need to model altruism explicitly.

The general conclusion is that when groups of people whose productivity is increasing relatively quickly have a great degree of influence on policy-making, we can expect the tax rate to decrease over time, and the government to accumulate assets. We arrive at the opposite prediction if the political power is more concentrated with the people whose productivity grows relatively slowly, in which case the expectation is of an increasing tax rate and the

---

<sup>37</sup> Altruism can be formally modelled by including the utility of other agents in each agent's utility function. If the overall utility of every agent is a weighted sum of her own intrinsic utility (from consumption and leisure), and the utility of other agents, the individually optimal policies with altruism are equivalent to the original specification without altruism, but with a positive Pareto weight on other agents.

<sup>38</sup> See for example Farhi and Werning (2007) for a discussion of this point in the context of bequest taxation.

government accumulating debt. These predictions are weakened by altruism between agents with different individually optimal policies.

## 2.6 Conclusion

Labour productivity trends of magnitude observed in many advanced economies in recent decades introduce a significant trend component into the optimal policy, and the policy becomes strongly sensitive to the choice of Pareto weights. For a range of Pareto weights, the model can explain the observed increases in debt in most industrial economies since the 1980s as the optimal policy response to an increase in labour income inequality.

In the horizon of decades, the change in policy induced by productivity trends usually quantitatively dominates the expected change in policy due to shocks. When agents face significantly heterogeneous productivity trends, which has been the case between 1980s and 2010s, a representative agent model, or a heterogeneous agents model without productivity trends yields quantitatively wrong conclusions about the optimal policy, because it neglects the conflict between agents over the timing of tax collection.

Apart from analysing past trends, this chapter presents a framework for the analysis of optimal policy in the future for various scenarios regarding the future development of productivity trends. In contrast, a standard model without productivity trends implicitly assumes that the labour productivity of all agents grows at the same rate. I discuss the policy-making implications in more detail in Chapter 3.

The strong dependence of the solution on the Pareto weights suggests the importance of political economics considerations for understanding observed policies and making testable predictions.

## Chapter 3

# Heterogeneous labour productivity trends and optimal fiscal policy with transfers

### 3.1 Introduction

Since the 1980s long-run trends such as technological change affected systematically differently the low-skilled and the high-skilled workers, and contributed to increasing earnings inequality and job polarisation across the majority of advanced economies (Acemoglu and Autor (2011)). In the previous chapter, I model these trends as heterogeneous productivity trends and show that optimal fiscal policy is significantly affected by their presence, as the *timing* of tax collection has distributional consequences.

In this chapter, I study the influence of heterogeneous productivity trends on the optimal policy in a different setting, in which the social planner can use uniform transfers in addition to the linear taxation of labour income, following Bhandari et al. (2017). I am chiefly motivated by two concerns.

Firstly, achieving redistribution by deviating from a flat tax profile is a rather blunt policy instrument. It is potentially very costly due to the convex distortionary cost of taxation (deviating from the 'tax-smoothing' outcome of Barro (1979)). Furthermore, not much

redistribution can be achieved because the agents endogeneously respond to any attempts of the policymaker to exploit the differences in the distribution of labour earnings over time, and these differences may not be very large in the first place unless there is a significant degree of heterogeneity in growth rates of productivity.<sup>1</sup> If transfers are available, the social planner may be able to achieve the majority of her distributional goals by using these, and refrain from the costly changes in the labour tax rate over time. In this sense, solving the model with transfers serves as a robustness check, as it shows which conclusions of the previous chapter depend on the assumption of the absence of transfers.

The second reason for introducing transfers is to make the analysis of the implications of productivity trends more relevant for policy-making considerations. Bhandari et al. (2017) argue that an affine tax function is a better approximation of the historically used tax policies in the US than a linear tax function. The model may offer a better reflection of the constraints faced by policymakers than the model without transfers. As I show in this chapter, the optimal policy is strongly affected by the addition of transfers. Therefore, it is important to understand the implications of productivity trends for the optimal policy in both settings.

As in the previous chapter, I study the optimal policy with two types of agents (high-skilled and low-skilled), when productivity trends of these types are calibrated to match the increasing earnings inequality between 1980s and 2010s in the US. For comparing the models, there are three useful conclusions. Firstly, in both models the change in the optimal policy due to trends in the horizon of decades is usually greater than the expected change due to expenditure shocks. Secondly, the trend component in the optimal tax rate is less pronounced than in the model without transfers for a wide range of Pareto weights, but more pronounced if the social planner cares very little about the welfare of the low-skilled. As shown by Bhandari et al. (2013), government debt is indeterminate in the model with transfers. Thirdly, the nature of the conflict between the agents over policy is fundamentally different. When the social planner uses transfers, the socially optimal policy strongly depends

---

<sup>1</sup>The statement that 'not much redistribution' can be achieved refers to the fact that the welfare cost of not being a dictator (not implementing one's individually optimal policy) is small in the model without transfers, as illustrated by Figure 2.8. In the limit case of no growth and no shocks, no redistribution through changing timing of taxation is possible, and the optimal policy entails a constant tax rate over time which is independent of Pareto weights.



on the Pareto weights even in the absence of productivity trends, as the low-skilled prefer a high level of labour taxation and transfers in all periods. Increasing inequality merely exacerbates this conflict. A choice of Pareto weights in the social planner's problem has much greater welfare implications than in the previous model.

The model developed in this article is well suited for making policy recommendations, if it is possible to predict the future evolution of labour productivities more accurately than assuming that the productivity of all types of workers will grow at the same rate, which is the implicit assumption of a model without productivity trends. While predicting the future is inherently difficult, there are reasons to believe that the assumption of a balanced growth of productivities may not be the best central prediction. To illustrate the quantitative effect of different predictions of future productivity trends on the optimal policy, I solve the model for a range of scenarios regarding the future evolution of productivity trends. The optimal policy in the horizon of decades is significantly affected by even modest heterogeneity in productivity trends.

All three articles in this thesis quantitatively study optimal fiscal policy in DSGE models with heterogeneous agents and incomplete markets. This chapter contains a discussion of the solution methods common to all articles. I explain the choice of the solution method, which is a parallel discrete value function iteration. The advantage of this method is its comparative stability. However, it suffers from problems related to the curse of dimensionality. I also implement a modification of the algorithm based on a parametric approximation of the value function (fitted value function iteration), which has the potential to overcome many of these problems. A learning algorithm is used for updating the coefficients which facilitates scalability, efficiency, and robustness to errors. Unfortunately, even though this approach is promising, the convergence properties of the resulting algorithm remain problematic.

This article is most closely related to Bhandari et al. (2013) and Bhandari et al. (2017),<sup>2</sup> with the addition of the heterogeneous productivity trends. In contrast, the previous chapter is more closely related to Shin (2006).

The rest of this chapter is organised as follows. Section 3.2 uses a two-period model to intuitively explore the implications of introducing transfers for the role of productivity trends in determination of the optimal policy. The full model is given in Section 3.3. Section 3.4 contains a quantitative analysis of the implications of heterogeneous productivity trends for optimal fiscal policy. Solution methods are discussed in Section 3.5. Section 3.6 concludes. Appendix B1 contains further definitions and results. Appendix B2 contains simulations of the solution for extreme choices of Pareto weights, which correspond to individually optimal policies of the agents. Appendix B3 is the main numerical appendix of this thesis and it is relevant also to Chapters 2 and 4.

## 3.2 Two-period model of optimal policy with transfers

As in the previous article I begin the presentation with a two-period simplified version of the model in order to intuitively explore the implications of productivity trends for the optimal policy. The introduction of transfers changes little in terms of the environment, but the characterisation of the optimal policy is markedly different.<sup>3</sup>

The government of a small open economy finances its exogenous expenditures by issuing debt and linearly taxing labour income. It can also use uniform transfers unrestricted in sign. The government solves a standard Ramsey problem, choosing a policy which maximises the social welfare assuming that all agents respond optimally to the policy. There are two periods indexed by  $t = 0, 1$ , and  $I$  agents indexed by  $i$ , with mass  $\pi_i > 0$ ,  $\sum_i \pi_i = 1$ . Agent  $i$ 's utility is  $U_i = u(c_{i,0}) + \beta u(c_{i,1})$ , where  $c_{i,t}$  is agent  $i$ 's consumption in period  $t$ ,  $u' > 0$ ,  $u'' < 0$ , and

<sup>2</sup>The 2013 paper is a much more extensive working paper version which contains additional details, such as a recursive formulation of the problem, and some features of the environment such as borrowing constraints are different. I refer to both papers depending on which one is more relevant in a particular context. Another closely related paper is Werning (2007), which studies a similar environment when markets are complete.

<sup>3</sup>The model is essentially the same as the two-period model of Chapter 2, with the exceptions of the budget constraints of the households and the government, and the set of policy instruments. Chapter 2 contains a more detailed discussion of the assumptions of the model, and of the relation to the full model.

$\beta \in (0, 1)$ . The production technology is CRS in labour supply. The agents differ in their initial productivity of labour and in its growth rate.

Agent  $i$ 's initial labour productivity is  $\theta_{i,0}$  and the growth rate of her productivity is  $\xi_i$ , so  $i$ 's productivity in the second period is  $\theta_{i,1} = (1 + \xi_i)\theta_{i,0}$ . All agents supply one unit of labour in each period so the aggregate product in period  $t$  is  $Y_t = \sum_i \pi_i \theta_{i,t}$ . I assume that there is a feasible tax policy, i.e. a tax policy which induces a competitive equilibrium in which the government's budget constraint is satisfied.

The agents and the government can save and borrow without any constraints at an exogenous gross interest rate  $R > 0$ .  $G > 0$  is the present value of the exogenous government expenditures. The government raises revenue by taxing labour income in the two periods at rates  $\tau_0, \tau_1$ . It can additionally use uniform transfers.  $T$  denotes per-capita transfers in the first period which can be negative, in which case they are interpreted as a lump-sum tax.<sup>4</sup> Taxation is associated with an exogenous distortion (collection costs) function  $D$ , so that distortion in period  $t$  for a given tax rate  $\tau_t$  and aggregate production  $Y_t$  is  $D(\tau_t, Y_t)$ , with  $\frac{\partial D}{\partial \tau} > 0, \frac{\partial^2 D}{\partial \tau^2} > 0$ . Transfers do not affect distortion. The government's budget constraint is

$$G + T \leq \tau_0 Y_0 - D(\tau_0, Y_0) + \frac{1}{R} (\tau_1 Y_1 - D(\tau_1, Y_1)), \quad (3.1)$$

and agent  $i$ 's budget constraint is

$$c_{i,0} + \frac{1}{R} c_{i,1} \leq (1 - \tau_0) \theta_{i,0} + \frac{1}{R} (1 - \tau_1) \theta_{i,1} + T. \quad (3.2)$$

Agent  $i$  chooses consumption  $c_{i,0}, c_{i,1}$  in the two periods to maximise her utility  $U_i$ , subject to her budget constraint (3.2), and taking the tax policy  $(\tau_0, \tau_1, T)$  as given. A competitive equilibrium given a tax policy  $(\tau_0, \tau_1, T)$  is a consumption allocation  $\{c_{i,t}\}_{i,t}$  which solves the utility maximisation problem of all agents, and the budget constraints of all agents and the government are satisfied.

<sup>4</sup>Because the interest rate is exogenous and common to all and there are no borrowing constraints, only the present value of transfers affects the equilibrium allocation. We can, without loss of generality, assume that the transfers are realised in the first period. Because the mass of agents sums to 1, so that  $\sum_i \pi_i T = T$ , the ' $T$ ' in the government's budget constraint is the same as in the constraints of the households.

The *socially optimal policy* (Ramsey plan) is a tax policy  $(\tau_0, \tau_1, T)$  for which a competitive equilibrium allocation results in the maximum possible social welfare  $\sum_i \alpha_i U_i$ , where  $\alpha_i \geq 0, i = 1, \dots, I$  are Pareto weights. In contrast to the model without transfers, the overall level of tax collection is not fixed, because it depends on endogenous transfers.

In the two-period model I focus on the individually optimal policies. The *individually optimal policy of agent  $i$*  is a socially optimal policy in which the Pareto weight is positive on agent  $i$  and zero on every other agent ( $\alpha_i > 0, \alpha_j = 0$  for  $i \neq j$ ). It can be shown that the individually optimal policy is a policy which maximises the right-hand side of agent  $i$ 's budget constraint (3.2), which is the present value of income of agent  $i$ , subject to the government's budget constraint (3.1). The first-order conditions for agent  $i$ 's individually optimal policies imply

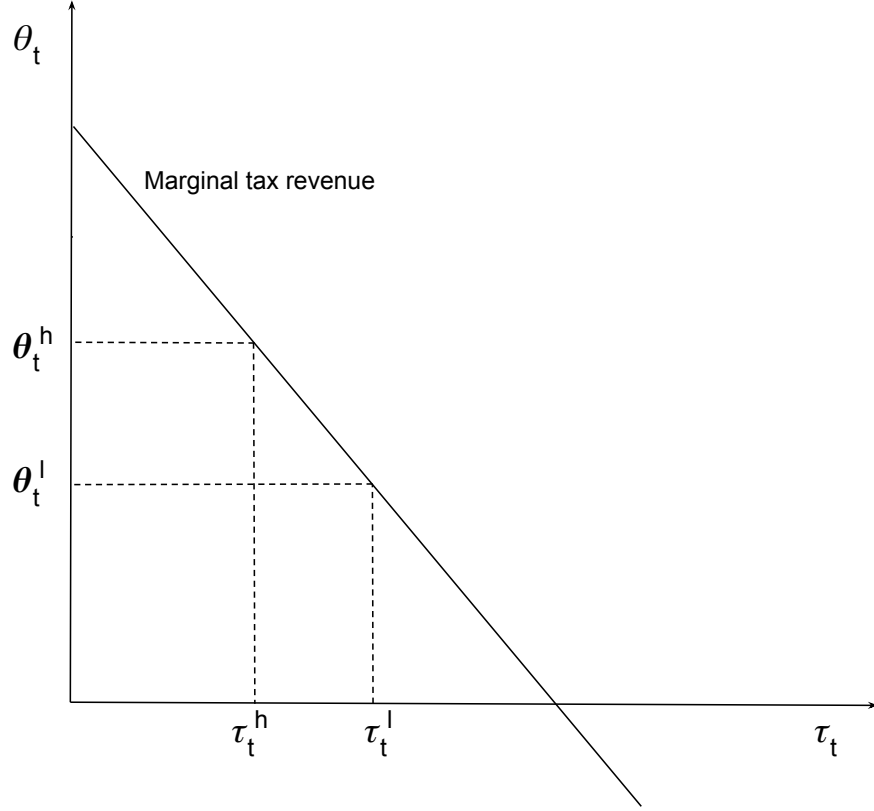
$$\theta_{i,t} = Y_t - \frac{\partial D(\tau_t, Y_t)}{\partial \tau_t} \equiv \text{MTR}_t, \quad (3.3)$$

$t=0,1$ . The interpretation is as follows. Consider the effect of a change in policy (starting from a feasible policy) by marginally increasing the tax rate in period  $t$ , and using the extra revenue to increase transfers  $T$ . From the perspective of agent  $i$  the marginal cost of such policy change is  $\frac{1}{R}\theta_{i,t}$ , which is the additional labour tax she pays. The marginal benefit is the (discounted) marginal tax revenue  $\frac{1}{R}\text{MTR}_t$ , as that is the increase in her after-tax income following from the increase in transfers. If a policy does not satisfy equation (3.3) it is not individually optimal for agent  $i$ , because we can find a feasible policy which yields higher after-tax income for agent  $i$ . In the model without transfers, the marginal benefit for an agent which follows from the increase in  $\tau_t$  always depends on her productivity in the other period, because the extra revenue can be used only to cut the tax rate on labour in the other period, and every agent benefits from this in proportion to her income in the other period.

Using equation (3.3), we can determine the individually optimal tax rate in each period in isolation from the other period. We can then find the individually optimal transfers using the government's budget constraint. Because distortion  $D$  is convex in  $\tau$  (so the marginal tax revenue is decreasing in  $\tau$ ) the agents with relatively high (low) productivity in a given period prefer a low (high) tax rate in that period. Figure 3.1 illustrates this point on the example

of a quadratic distortion function. Figure 3.2 shows the optimal tax rate  $\tau_t$  in a period as a decreasing function of the agent's productivity.<sup>5</sup>

Fig. 3.1 Individually optimal tax rate for high ( $\theta_t = \theta_t^h$ ) and low ( $\theta_t = \theta_t^l$ ) productivity agent



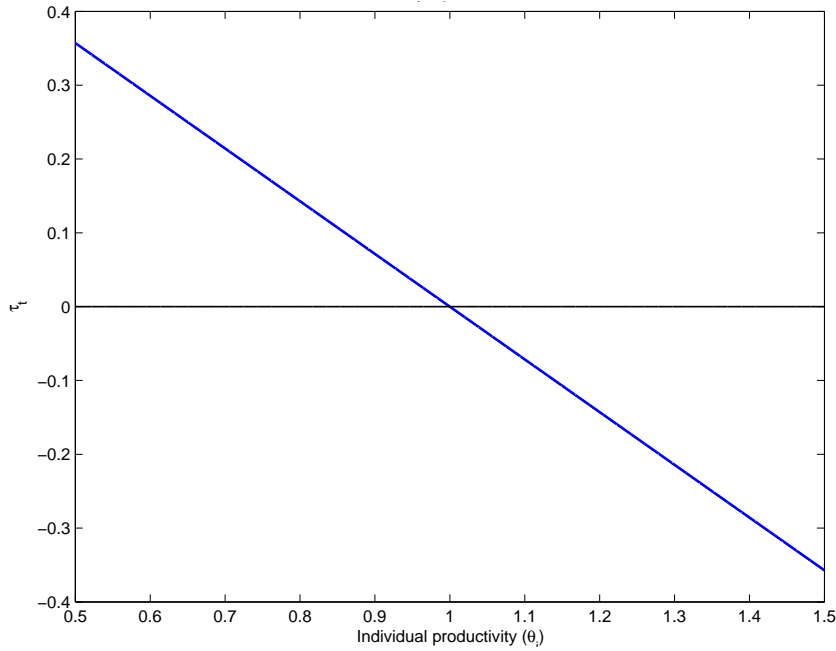
Dividing the period 1 version of (3.3) by the period 0 version we obtain the same optimality condition as in the model without transfers

$$1 + \xi_i = \frac{\theta_{i,1}}{\theta_{i,0}} = \frac{Y_1 - \frac{\partial D(\tau_1, Y_1)}{\partial \tau_1}}{Y_0 - \frac{\partial D(\tau_0, Y_0)}{\partial \tau_0}}. \quad (3.4)$$

As in the model without transfers, the agents with a higher (lower) growth rate of productivity  $\xi_i$  prefer lower (higher) taxes in the future. However, the initial level of productivity now

<sup>5</sup>In this example the individually optimal policy of agents who are highly productive within a period actually entails a negative tax rate (labour subsidy). The quadratic distortion is symmetric around zero, and in the figure the MTR is extended for negative values of tax rate. This can be thought of as labour subsidy expanding the tax base, so the government has to pay out less in terms of subsidies if a negative tax rate moves closer to zero (and the 'distortion' is thus lower).

Fig. 3.2 Individually optimal tax rate (with transfers)



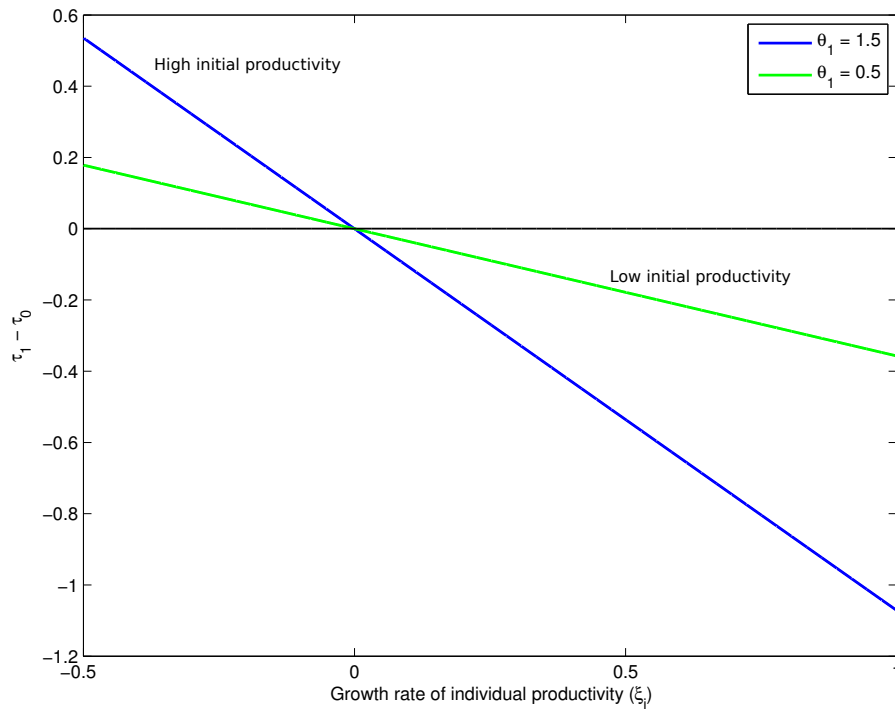
affects the change in the tax rate over time, as the level of productivity in each period fixes the tax rate in that period (equation (3.3)). For a given growth rate of productivity  $\xi_i$ , agents with greater initial productivity prefer a greater change in tax rate over time, because the change in their level of productivity is greater. Figure 3.3 shows the increase in agent  $i$ 's individually optimal tax rate as a function of her growth rate of productivity, for two initial levels of productivity for the quadratic distortion example.<sup>6</sup> In the corresponding figure (2.3) in the model without transfers the slope was unaffected by the initial level of productivity.

We can make two broad conclusions. Firstly, with regard to the individually optimal policies both the heterogeneity in levels and in growth rates now matters. Even if the income of two agents grows at the same rate, their individually optimal policy is different if they differ in the initial level of productivity. The interaction of the distribution of levels and growth rates is of crucial importance in finding the optimal policy.

Secondly, in the model with transfers a distributional conflict is present already in a static or balanced growth environment. Even if everyone's productivity grows at the same rate,

<sup>6</sup>Both lines cross at  $\xi_i = 0$  because the growth of the aggregate product was assumed to be zero in constructing this example, so for  $\xi_i = 0$  equation (3.3) has the same solution in both periods (and  $\tau_t$  is constant over time).

Fig. 3.3 Increase in individually optimal tax rate (with transfers)



agents with different initial level of productivity prefer a different tax policy and the solution thus depends on the preferences of the social planner. In the model without transfers the heterogeneity in growth creates a conflict. In the model with transfers it merely changes an existing conflict (and exacerbates it if inequality is increasing over time).

Another useful way of thinking about the differences between the two models comes from interpreting the Ramsey problem in the model with transfers as a two-stage problem. In the first stage, the socially optimal present value of transfers  $T^*$  is determined, which depends both on the distribution of the initial productivities and the distribution of growth rates of productivity. In the second stage the timing of tax collection is decided, i.e. how much of the tax revenue of present value  $T^* + G$  should be collected in each period. The second stage is the same as the Ramsey problem in the model without transfers, in which we also determine the optimal timing of tax collection for a given revenue. The results discussed in the previous chapter apply to the second stage of the problem. Once a level of transfers is set, all agents with the same growth rate of productivity prefer the same timing of tax collection.

### 3.3 The full model

In this section I set up a model of optimal policy in an economy in which agents face heterogeneous productivity trends. The social planner can use uniform transfers as in Bhandari et al. (2013), in addition to linearly taxing labour earnings. The model is very close to the model of Chapter 2, the only substantial differences being that the budget constraints and the set of policy instruments now contain transfers. A reader familiar with the model of the previous chapter may wish to inspect the budget constraints (equations (3.10), (3.11)) and then proceed directly to Subsection 3.3.1, where I address the implications of introducing transfers for the equilibrium and the optimal policy.<sup>7</sup>

#### The environment

There are  $I$  types of infinitely lived agents indexed by  $i = 1, \dots, I$ . The mass of agent of type  $i$  is  $\pi_i$ , with  $\sum_{i=1}^I \pi_i = 1$ . Agent  $i$ 's objective function is defined over stochastic sequences of consumption  $\{c_{i,t}\}_{t=0}^{\infty}$  and labour  $\{l_{i,t}\}_{t=0}^{\infty}$ :

$$\mathbb{E}_0 \left[ \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right]. \quad (3.5)$$

$U^i$  is the instantaneous utility function,  $\beta \in (0, 1)$  is the discount factor, and  $\mathbb{E}_t$  denotes expectation conditional on information available in period  $t$ . Labour  $l_{i,t} \in [0, \bar{l}_i]$ , where  $\bar{l}_i > 0$  is given. The utility functions are twice continuously differentiable with  $U_c^i > 0, U_l^i < 0, U_{cc}^i < 0, U_{ll}^i < 0$ .

One unit of labour supplied by type  $i$  agent in period  $t$  produces  $\theta_{i,t}$  units of output, and the aggregate resource (feasibility) constraint is

$$\sum_{i=1}^I \pi_i c_{i,t} + g_t \leq \sum_{i=1}^I \pi_i \theta_{i,t} l_{i,t} \quad (3.6)$$

---

<sup>7</sup>An alternative approach would be to only state in the main text the parts which are different from the previous chapter, and put the full statement of the model in an appendix. However, since this thesis is of the three article format rather than the book format I believe that it is good practice to keep the individual articles relatively self-contained. The statement of the model is kept as brief as possible and standard definitions are relegated to Appendix B1.



for  $t = 0, 1, \dots$ , where  $g_t$  is government expenditure at  $t$ .

The adjusted government expenditure  $\{\bar{g}_t\}_{t=0}^\infty$  is

$$\bar{g}_t = k \sum_{i=1}^I \pi_i \theta_{i,t} \bar{l}_i \quad (3.7)$$

for  $t = 0, 1, \dots$  and some given  $k > 0$ . The actual government expenditure in period  $t$  is

$$g_t = s_t \bar{g}_t \quad (3.8)$$

for  $t = 0, 1, \dots$ , where  $s_t$  follows a finite-state Markov process, with  $s_t > 0$  for all  $t$ , and transition probabilities are denoted as  $Pr(s_{t+1}|s_t)$ .  $s^t = (s_0, \dots, s_t)$  denotes a history of shocks up to period  $t$ . Agent  $i$ 's labour productivity evolves according to an exogenous trend function

$$\theta_{i,t} = \theta_i(t), i = 1, \dots, I. \quad (3.9)$$

Functions  $\theta_i(t)$  become constant in  $t$  after  $\mathcal{T}$  periods, where  $\mathcal{T}$  can be large but finite.

Markets are incomplete. The only asset in the economy is a one-period risk-free bond issued by the government or the agents.  $b_{i,t}$  denotes holdings of this bond by agent  $i$  bought in period  $t$ , maturing in period  $t + 1$ . Let  $R_t$  denote a gross interest rate from time  $t$  to  $t + 1$ . Labour income of every agent in period  $t$  is taxed at a proportional tax rate  $\tau_t$ , and the agent receives per-capita transfers  $T_t$  which are not agent-specific. Transfers are not ex-ante restricted in sign.<sup>8</sup> Agent  $i$ 's sequence of budget constraints is

$$c_{i,t} + b_{i,t} \leq (1 - \tau_t) \theta_{i,t} l_{i,t} + R_{t-1} b_{i,t-1} + T_t, \quad (3.10)$$

for  $t = 0, 1, \dots$ . Without loss of generality we can assume that  $R_{-1} = 1$ .

---

<sup>8</sup>The Ramsey planner will never choose negative transfers (lump-sum taxes) except for cases of Pareto weights strongly favouring the highly productive agents, because these are very harmful to the agent with the lowest level of productivity. Therefore lump-sum taxes are not used to circumvent the trade-offs involving distortionary taxation.

Let  $B_t$  denote the government's assets purchased in period  $t$ , maturing in period  $t + 1$  ( $-B_t$  is the government's debt). The government's budget constraint is<sup>9</sup>

$$g_t + B_t + T_t \leq \tau_t \sum_{i=1}^I \pi_i \theta_{i,t} l_{i,t} + R_{t-1} B_{t-1}, \quad (3.11)$$

for  $t = 0, 1, \dots$

The agents and the government start with initial holdings of assets  $\{b_{i,-1}\}_{i=1}^I$ ,  $B_{-1}$ , and the market-clearing condition in every period  $t \geq -1$  is

$$\sum_{i=1}^I \pi_i b_{i,t} + B_t = 0. \quad (3.12)$$

Following Bhandari et al. (2013) I assume that the asset holdings of the agents and the government are bounded from below, rather than assuming a particular borrowing constraint. Due to market-clearing constraint, this implies that all assets are also bounded from above.

The social planner's objective function is

$$\mathbb{E}_0 \left[ \sum_{i=1}^I \pi_i \alpha_i \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right], \quad (3.13)$$

where  $\alpha_i > 0$  is Pareto-weight on agent  $i$  and  $\sum_i \alpha_i = 1$ .

### 3.3.1 Competitive equilibrium and the optimal policy

The definitions related to the competitive equilibrium and a socially optimal competitive equilibrium (Ramsey plan) are almost unchanged relative to the model without transfers and are relegated to Appendix B1. The only substantial differences are that the tax policy is now a pair  $\{\tau_t, T_t\}_t$  rather than just  $\{\tau_t\}_t$ , and the definitions are modified accordingly to take this into account.

When the social planner can use unrestricted transfers, the differences in asset holdings between agents play an important role in relation to equilibrium allocations. Following

---

<sup>9</sup>Because the total mass of agents equals 1 we can write  $T_t$  in the budget constraint of the government, as opposed to  $\sum_i \pi_i T_t$ .

Bhandari et al. (2013), I define the *net* asset position of agent  $i$  (without loss of generality) relative to agent 1 as  $\tilde{b}_i \equiv b_i - b_1, i = 2, \dots, I$ . The term *gross* asset position refers to the asset holdings in levels  $(b_1, \dots, b_I)$ .

The introduction of transfers has two important implications which carry over from Bhandari et al. (2013) and are stated here without proof.<sup>10</sup>

Firstly, in a competitive equilibrium only the net asset positions are determined, while the gross asset positions are indeterminate. For any competitive equilibrium (given a policy), if we take the equilibrium allocation and adjust the asset profile of all agents in such a way that the net asset positions are unchanged, we can find a government policy (by changing transfers and government assets while keeping the tax rate the same) which implements the original equilibrium allocation, along with the new asset profile.

The intuition for this result can be demonstrated in a steady state setting. Suppose that the economy is in a deterministic steady state in which the interest rate is  $R$ . If the asset holdings of all agents increase by the amount  $k$ , each agent receives additional revenue of  $(R - 1)k$  each period (after the repurchase of bonds). The value of government's debt increases by  $k$  so that markets clear, which means that it needs to collect  $(R - 1)k$  additional tax revenue every period if transfers are not adjusted. If transfers are instead decreased by  $(R - 1)k$  and the tax rate is kept as before, then every agent has the same amount of resources after paying taxes, receiving transfers, and purchasing bonds as in the original steady state and faces the same prices and tax rate, so chooses the original allocation. If the asset holdings change by a different amount for different agents so the net asset positions are not preserved, a compensating change in transfers cannot implement the original steady state.

The second important implication of introducing transfers is that the level of government debt is indeterminate. This is a consequence of the indeterminacy of gross asset positions, and the fact that the government debt is the sum of the gross asset positions of all agents due to the market-clearing condition. Without loss of generality, we can normalise by setting the government's debt to zero,  $B_t = 0, t = 0, 1, \dots$

<sup>10</sup>The relevant results are Theorem 1 and Corollary 1. The dependence of productivities on time does not change the validity of the proofs.

A useful consequence of the fact that only the net asset positions are determined in equilibrium is that we can find a recursive formulation of the problem with one fewer state variable than in the absence of transfers. This is discussed in more detail in Appendix B1.

### 3.4 Quantitative analysis of the optimal policy

In this section I use the model to study the implications of heterogeneous productivity trends in two settings. Firstly, I analyse the consequences of the increase in earnings inequality in the last several decades in the US for the optimal policy. Secondly, I study the policy-making implications of future productivity trends. In both settings the optimal policy is significantly affected by the presence of productivity trends.

#### 3.4.1 Increase in earnings inequality since the 1980s

In the first article, I show that in the absence of transfers, the optimal policy is significantly affected by productivity trends consistent with data between 1980 and 2010 in the US if the social planner does not use transfers. In this section I solve the optimal policy problem with transfers in the same setting. Studying the same environment demonstrates how the introduction of transfers changes the optimal policy in the presence of productivity trends, and the quantitative implications of trends in a model with transfers more broadly.

In the previous chapter, there were three important conclusions which I focus on for the purpose of comparing the models. Firstly, the expected change in policy due to trends was usually (for the majority of choices of Pareto weights) greater than the expected change in policy due to the stochastic nature of government expenditures. Secondly, it was usually optimal to set an increasing time path of taxation, which was accompanied by an increase in government debt. Thirdly, even though the optimal policy was very sensitive to Pareto weights, the welfare of the agents was not.

The first conclusion regarding the relative importance of shocks and trends stands. The second conclusion is different. The time path of taxation is still usually increasing, but the magnitude of the change is smaller, unless the social planner places a very low weight on the

welfare of the low-skilled. Because the government debt is indeterminate, an increasing tax rate over time is no longer associated with an increase in government debt. Finally, the choice of Pareto weights when transfers are available has much greater welfare consequences.

The calibration is the same as in the previous chapter. There are two agents of equal mass, and the evolution of their productivity follows the 1980 - 2010 evolution of 90-10 earnings ratio in the US. Agent 1 ('high-skilled') is initially more productive and her productivity grows at a constant rate. Agent 2 ('low-skilled') is less productive, and her productivity is fixed at the initial level. Further calibration details are explained in the previous chapter and in Appendix A2.

In order to study the effect on the optimal policy of productivity trends in isolation from the effect of shocks, I solve the model for a stochastic calibration with productivities fixed at their initial level, and for a deterministic calibration with growth.

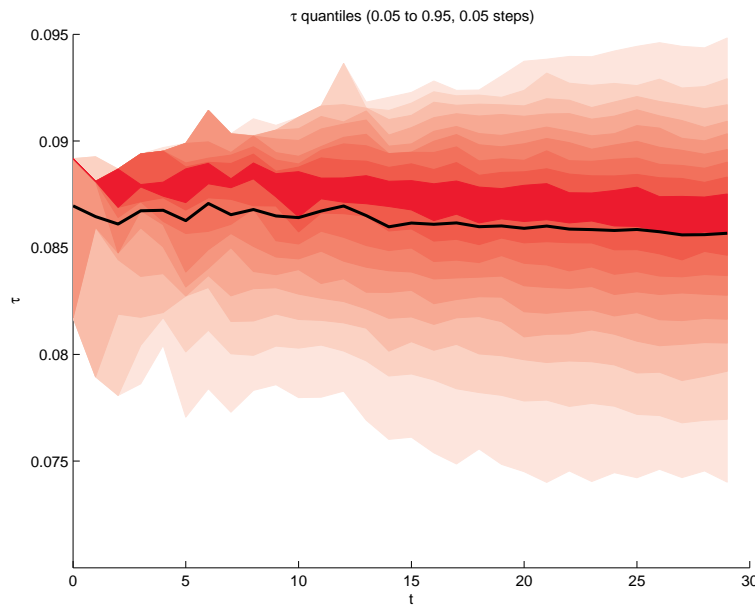
Figure 3.4 shows the effect of shocks on the tax rate in a calibration without growth, and Pareto weights<sup>11</sup>  $\alpha_1 = 0.8$ ,  $\alpha_2 = 0.2$ . Every period sample quantiles of the tax rate from 1,000 simulations are shown, along with the sample mean. The tax rate responds to shocks significantly, and it is relatively high in times of high expenditures. However, the expected change in tax rate over 30 years is quantitatively small (approximately 0.001). For different choices of Pareto weights the fluctuations in the tax rate are different, but the result of a small *expected* change in policy over a period of several decades is unaltered. A deterministic calibration with growth is therefore likely to yield very similar results to a stochastic calibration with respect to the expected change in policy, which is the focus of my analysis,<sup>12</sup> and the computation time is significantly reduced.

Figure 3.5 shows the solution for a deterministic calibration ( $s_t \equiv 1$ ) with equal Pareto weights on both agents. Because the labour productivity of the low-skilled is lower in all periods, which implies a higher marginal utility of consumption in the absence of redistribution, the social planner chooses a policy which redistributes resources to the low-skilled. This is

<sup>11</sup>The choice of these weights is motivated by the fact that (as I show later) in this instance the productivity trends result in a small change in policy over time. It is thus a case in which the effect of shocks on policy relative to the effect of trends could be relatively significant.

<sup>12</sup>I verify this numerically for the case of equal Pareto weights. Between the stochastic and the deterministic calibrations, the difference in the expected tax rate after 30 years is of the order of  $10^{-4}$ .

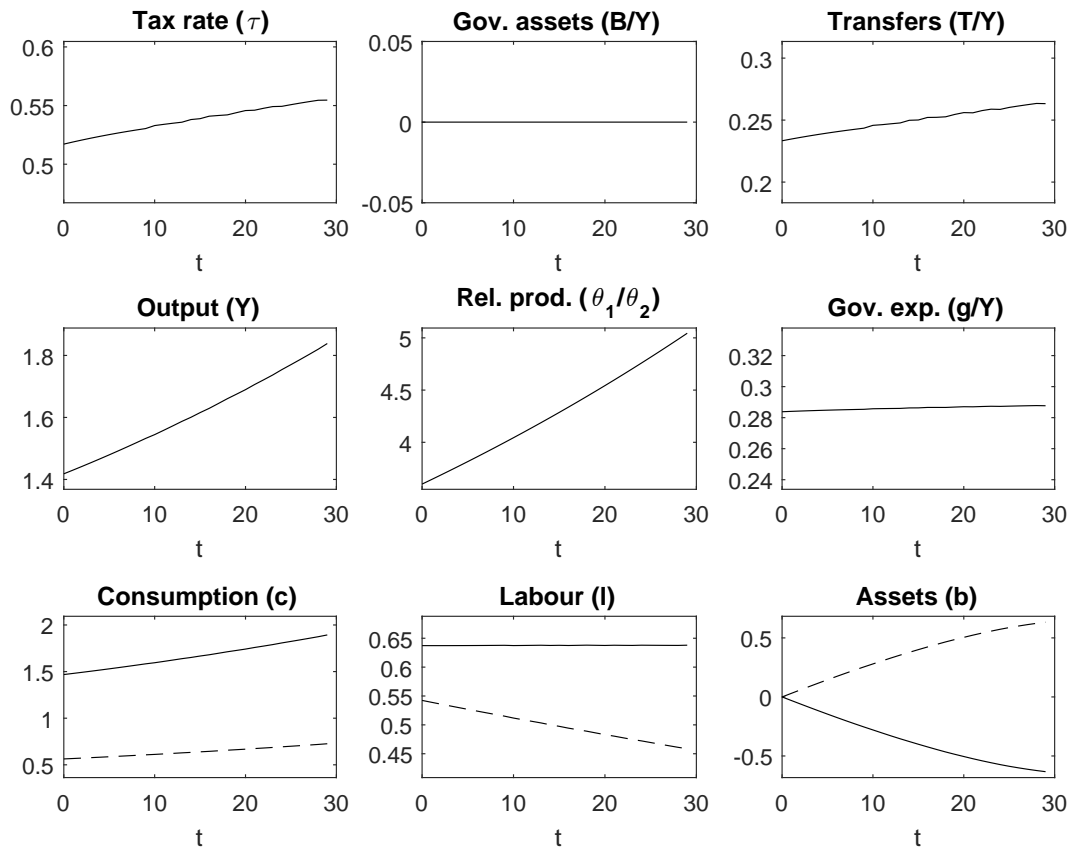
Fig. 3.4 Effect of shocks on the tax policy in a static environment (no growth)



implemented by taxing labour at a higher rate than the financing of government expenditures requires, and using the extra revenue to finance uniform transfers. In later periods there is a greater labour income inequality due to diverging productivities, and the tax rate rises by 4 percentage points over 30 years, financing higher transfers. This is consistent with the two-period model which suggests that the agent whose productivity grows at a lower rate benefits from an increasing tax rate over time. The government debt is normalised to zero. Appendix B2 contains the equivalents of Figure 3.5 for more extreme choices of Pareto weights which correspond to the individually optimal policies studied in the two-period model.

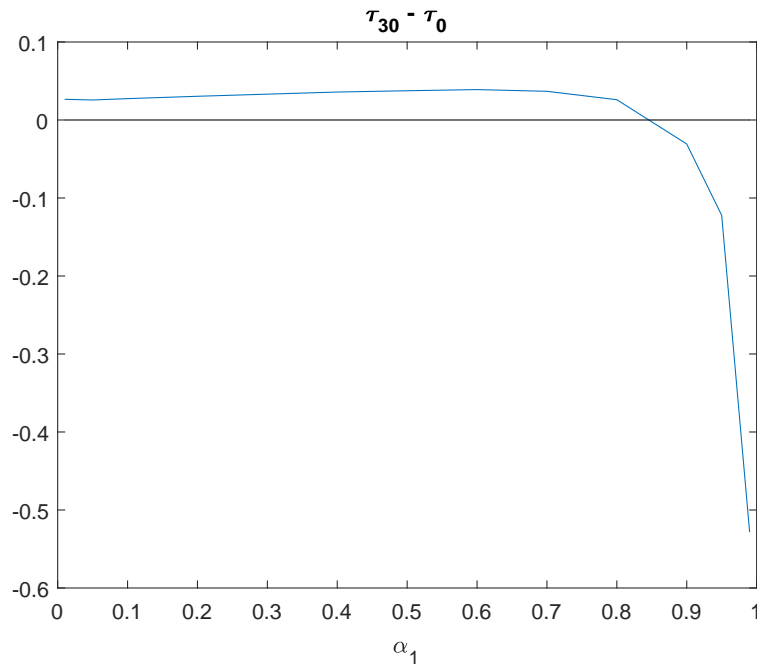
Figure 3.6 shows the optimal increase in the tax rate over a 30-year period starting in 1980 as a function of the Pareto weight on the high-skilled agent, and Figure 3.7 shows the corresponding increase in transfers. As in the model without transfers, the productivity trends imply increasing tax rate for the majority of Pareto weights but the change in the tax rate is usually smaller, except for cases when the social planner places a value on the welfare of the high-skilled of approximately  $\alpha_1 = 0.9$  or higher. The low-skilled still benefit from postponing taxation. For  $\alpha_1 = 0$ , the tax rate is already very high in the first period (72%). Therefore, increasing it further is very costly in terms of distortion, and the change over time

Fig. 3.5 Optimal policy over 30 years without shocks ( $\alpha_1 = \alpha_2 = 0.5$ ). In the last row the solid lines correspond to the high-skilled agent (1), the dashed lines to the low-skilled (2).



is small (3 percentage points). As  $\alpha_1$  increases, the Ramsey planner wants to redistribute to the low-skilled to a lesser extent, but doing so by increasing the tax rate over time is less costly because the distortion is lower in the first period. The overall effect of an increase in  $\alpha_1$  on the change in the tax rate is unclear, unlike in the case of no transfers where the increase in the tax rate ( $\tau_{30} - \tau_0$ ) was a decreasing function of  $\alpha_1$ . The increase in the tax rate is slightly positive and close to constant for  $\alpha_1$  ranging from zero to 0.75 (the tax rate increases by 2-4 percentage points).

Fig. 3.6 Increase in the tax rate over 30 years

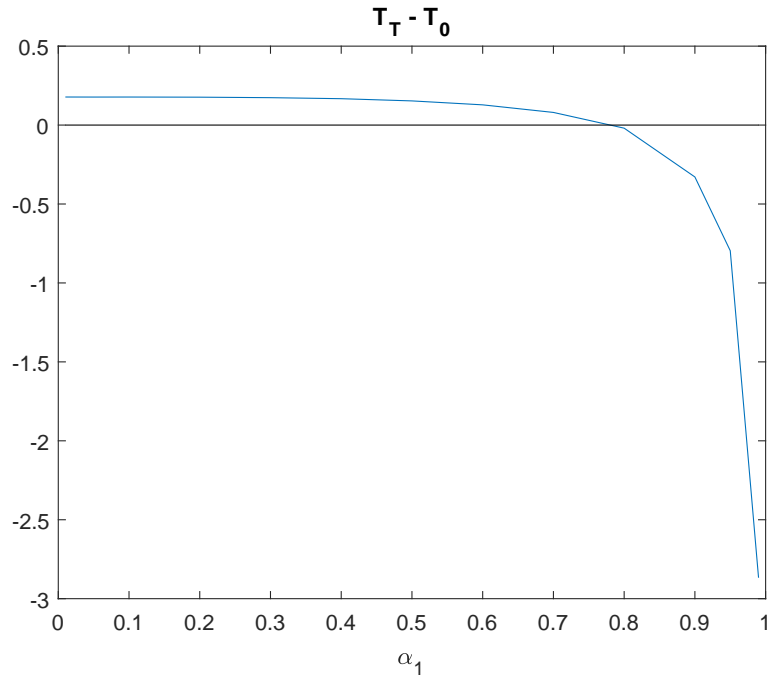


For  $\alpha_1$  close to 0.85 an approximately constant tax rate over time is optimal. For greater  $\alpha_1$  the social planner chooses a policy which favours the high-skilled, which amounts to a decreasing tax rate. As  $\alpha_1$  increases further the distortion is weaker and the absolute value of the change increases. For  $\alpha_1$  very close to 1, the social planner actually wants to use lump sum taxes. The revenue is used to finance a labour subsidy which benefits the high-skilled disproportionately. In this case a decrease in  $\tau$  represents an increase in labour subsidy.

The change in the tax rate due to trends is still quantitatively significant but usually less pronounced than in the model without transfers. However, the expected change in policy due to shocks is also smaller, as was demonstrated using the static calibration. The conclusion



Fig. 3.7 Increase in transfers over 30 years



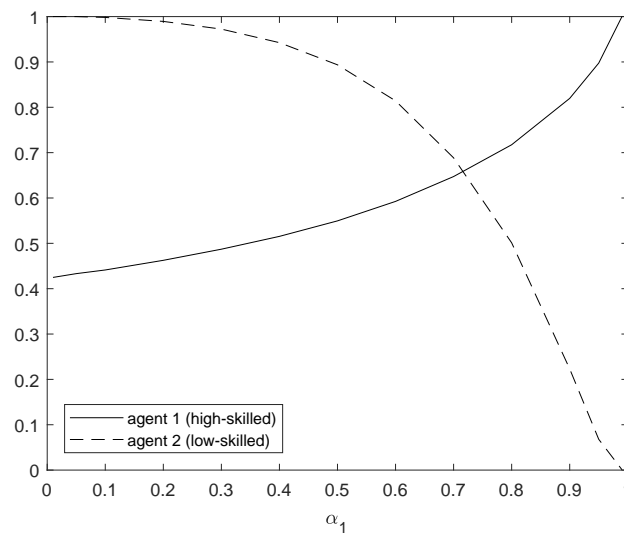
that heterogeneous productivity trends have a greater impact on the expected change in policy than expenditure shocks in the horizon of decades is thus unaltered by the introduction of transfers.

The optimal policy strongly depends on Pareto weights, as does the welfare of the individual agents. Figure 3.8 is a counterpart of Figure 2.8 in the model without transfers. It shows, for every agent, the welfare cost of moving from the agent's individually optimal equilibrium to a socially optimal equilibrium with Pareto weight  $\alpha_1$  on the high-skilled and  $\alpha_2 = 1 - \alpha_1$  on the low-skilled. It is expressed as a consumption equivalence. A value  $k$  means that the allocation is equivalent in terms of utility to the individually optimal allocation, if its consumption element is multiplied by  $k$ . For example, the value of 0.42 for the high-skilled when  $\alpha_1 = 0$  means that if the social planner chooses the individually optimal policy of the low-skilled, the high-skilled are as well off as if the individually optimal policy of the high-skilled was chosen, and they subsequently lost 58 percent of their consumption.

For the low-skilled, the welfare consequences of the individually optimal policy of the high-skilled can be especially dire, equivalent to losing almost all of their consumption

starting from their individually optimal policy. The individually optimal policy of the high-skilled entails lump-sum taxes which force the low-skilled to supply a great amount of labour to satisfy their budget constraint, and which are used for labour subsidies that benefit primarily the high-skilled. The high-skilled, in contrast, cannot be immiserated. These results are in stark contrast to the model without transfers in which the Pareto weights strongly affected policy but not individual welfare.<sup>13</sup>

Fig. 3.8 Consumption equivalence relative to the individually optimal policy



In the model with transfers the Pareto weights are important for welfare even in the absence of increasing inequality. An equivalent of Figure 3.8 obtained in a static environment without growth is similar, but the consumption equivalence curves are slightly less steep moving away from the individual optimum, reflecting the fact that the presence of growth makes the conflict over policies stronger. The cost for the high-skilled of moving from their individually optimal policy to the individually optimal policy of the low-skilled decreases from 58 percent of consumption to 50 percent.

<sup>13</sup>For either the high-skilled or the low-skilled, the decrease in utility associated with moving from their individually optimal policy to the individually optimal policy of the other agent was equivalent to losing no more than 2 percent of consumption.

### 3.4.2 Policy implications of future productivity trends

The analysis so far focused on past productivity trends. The model is also useful for a study of policy implications of future evolution of productivity trends. The implicit assumption of conventional optimal policy models is that of a balanced growth of labour productivities of all agents. If we are able to predict the future distribution of labour productivity more accurately than assuming balanced growth, then using a conventional model amounts to discarding valuable information and yields potentially suboptimal policy recommendations.

Predicting the future is inherently difficult but there are two reasons to believe that the assumption of balanced growth may not be the best central prediction. The first is based on extrapolation of recent trends, the other is connected to the role of technological change.

As Chapter 2 discussed, there was a significant heterogeneity in labour productivity trends in many advanced economies in the last several decades. From the perspective of the present, the assumption of balanced growth of labour productivities implies a discontinuous adjustment in the growth rates of productivities of different types of workers (high skilled and low skilled) to the same level. Even if we expect little increase in inequality in the long run, it is perhaps more appropriate to model this scenario as a gradual adjustment in the growth rates of different agents.

One of the causes of the increasingly unequal distribution of labour productivity in recent decades has been technological progress.<sup>14</sup> Brynjolfsson and McAfee (2014) argue that in recent years the progress in digital technologies in particular seems to be accelerating, and this is likely to cause a significant increase in inequality in the future. There is considerable uncertainty about the severity and timing of the technological disruption, but many academics and policy-makers believe that a continuing increase in inequality is likely.<sup>15</sup>

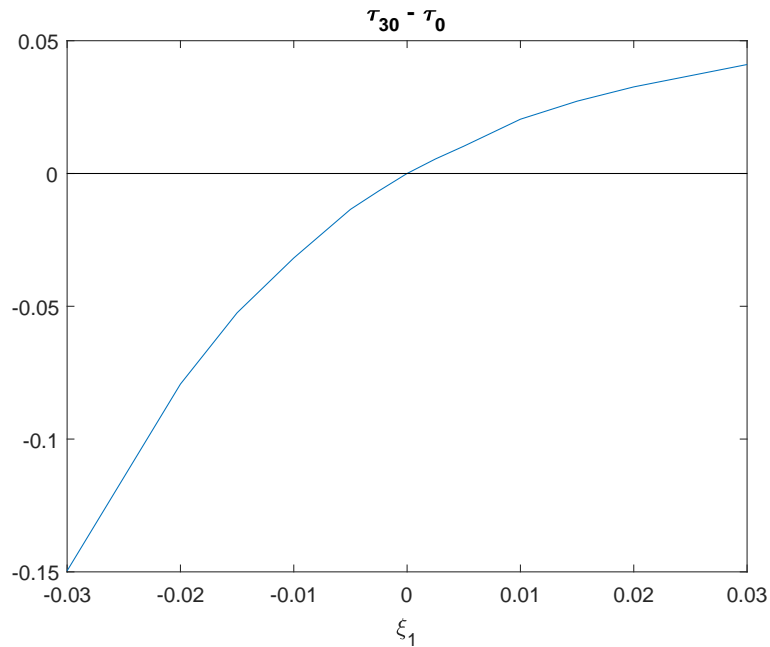
---

<sup>14</sup>Autor (2014) summarises empirical evidence for US data. Acemoglu and Restrepo (2017) focus on the role of industrial robots in depressing wages and increasing unemployment. In general, it is difficult to disentangle the effect of technological progress from other trends going on in the last several decades, but it appears that technological progress was a substantial force contributing to the rise in inequality.

<sup>15</sup>See Autor (2015a) for a recent summary of the debate. Autor (2015b) conjectures that the impact of technological change on earnings distribution may not be quantitatively large, and may be relatively short-lived due to complementarity effects between labour and technology, and due to labour supply response to technological change in the long run. Acemoglu and Restrepo (2016) build a model of technological progress, in which an increase in wage inequality may occur only in a transitory period rather than in the very long run.

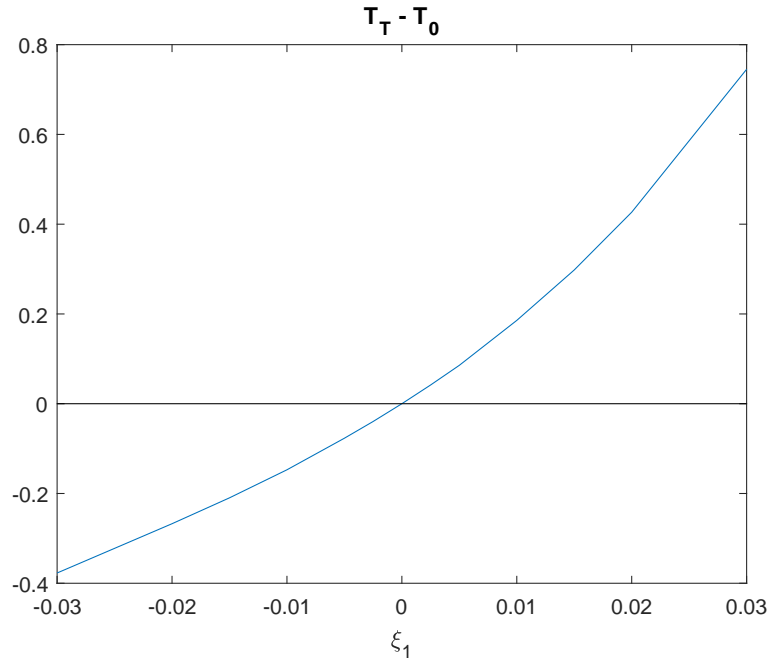
In order to illustrate the quantitative effect of different predictions of future productivity trends on the optimal policy, I solve the model for a calibration with two agents (high-skilled and low-skilled) as in the previous subsection. The calibration is essentially a continuation of the 2010 state of the US economy, with the initial productivities corresponding to the productivities in 2010. I assume that the growth rate of productivity of each agent is constant for the next 30 years, and I solve the model for a range of differences between the growth rate of the high-skilled ( $\xi_1$ ) and of the low-skilled (normalised to 0). There are no shocks.<sup>16</sup> The Pareto weight is equal on both types. Figures 3.9 and 3.10 show the optimal increase in the tax rate and transfers over a 30-year period starting in 2010, for a range of growth rate differences. Negative values of the difference ( $\xi_1$ ) mean that the inequality is decreasing. Unless the growth difference is close to zero, the optimal tax rate changes by at least several percentage points, accompanied by a change in transfers. In particular, if the growth rate difference remains the same as in the previous thirty years, the tax rate should increase by 2.5 percentage points.

Fig. 3.9 Optimal increase in the tax rate as a function of the difference in growth rates ( $\alpha_1 = \alpha_2$ )



<sup>16</sup>As discussed in the context of the analysis of past productivity trends, the shocks in the model with transfers introduce an expected change in the tax rate over a 30-year period that is quantitatively negligible.

Fig. 3.10 Optimal increase in transfers as a function of the difference in growth rates ( $\alpha_1 = \alpha_2$ )



## 3.5 Solution methods

In this section I give a high-level non-technical overview of the solution methods employed for obtaining the numerical results in this thesis. I discuss the rationale behind my choice of the solution method (discrete value function iteration), its limitations, and ways to overcome them.

This discussion refers mainly to Chapters 2 and 3, which study the optimal policy in the presence of heterogeneous growth. Chapter 4 contains a model of optimal policy in an open economy setting. The model is solved similarly but there are important differences which are addressed in the following chapter. A more technical discussion of implementation details is relegated to Appendix B3.

### 3.5.1 Overview of solution methods

The problems studied in this thesis require the use of a global solution method. The presence of heterogeneous productivity trends in Chapter 2 and 3 induces a significant variation over

time in important features of the environment (production technology) and states. A local approximation around a point in state space would provide an inaccurate approximation of the solution during the period of productivity growth which is the main focus of my analysis.<sup>17</sup> In the remainder of this section I focus on global solution methods.

As discussed earlier, I use numerical dynamic programming (NDP) to solve the models. In the context of solving optimal policy models with incomplete markets, this approach is usually applied to problems in which there are at most three continuous state variables (Shin (2006), Farhi (2010)). Optimal policy problems of higher dimension are commonly solved by different methods such as parametrised expectations (Den Haan and Marcet (1990)) or other projection methods.<sup>18</sup> The main issue is that convergence of these methods is not guaranteed, and usually depends on the initial guess. While numerical dynamic programming is also not guaranteed to converge, in practice it is often more reliable, and it is less sensitive to a choice of initial guess.<sup>19</sup>

The NDP solution approach relies on using a recursive formulation of the problem. Having settled on a recursive formulation, there are a number of choices to be made and their combination results in a vast number of possible algorithms.<sup>20</sup> Firstly, we need to choose which object to approximate, which is closely related to the basic structure of the algorithm. I choose to approximate the value function in periods  $t \geq 1$  and to use value function iteration (VFI).

Secondly, we need to choose how to approximate the value function. This is crucially important for computation time, accuracy of the solution, and convergence. The approximation method can be classified as either discrete (also known as finite element approximation), or parametric. The numerical results in this thesis were obtained using a discrete approximation.

<sup>17</sup>Kollmann et al. (2011) compare the accuracy of local and global solution methods in a different setting (a multi-country RBC model) and find that the accuracy of local approximations deteriorates rather quickly as we move away from the steady state of the model.

<sup>18</sup>See Judd (1998), or Heer and Maussner (2009) for an introduction to parametrised expectations approach. Faraglia et al. (2014b) discuss more recent developments and applications.

<sup>19</sup>The limited influence of the initial guess is particularly important for the analysis in this article, because the optimal policy is strongly affected by the calibration of productivity trends, and Pareto weights. Even if a solution was found in one setting, it would not necessarily be a good initial guess for a different setting.

<sup>20</sup>The book of Bertsekas (2012) provides an extensive survey of the field. Cai and Judd (2014) survey recent applications in economics.

### 3.5.2 Discrete approximation of the value function

A discrete approximation means that the value function is represented as a list of numbers, one corresponding to each point on a multi-dimensional discrete grid. States are not constrained to coincide with grid points and continuation values in between these are computed using interpolation. I do not interpolate over the discrete shock realisations, thus having effectively a separate value function for each shock realisation. This method has been used to solve optimal policy models with heterogeneous agents by Shin (2006), Farhi (2010), or Golosov and Sargent (2012).

The main advantage of a discrete value function iteration is that it is comparatively stable. However, it suffers from other problems, most of which are related to the curse of dimensionality, which means that the number of grid points increases exponentially with the number of states.<sup>21</sup> With four continuous states (as in Chapter 2) computing an accurate solution is very time-consuming. To mitigate the issue of computation time I parallelise the solution algorithm.<sup>22</sup>

At the beginning of iteration  $r$  of the VFI algorithm, we have a discrete approximation  $V_r$  of the value function. In the maximisation step we need to obtain  $V_{r+1}$ , which is a list of optimal values of the problem on the right-hand side of the Bellman equation at every grid point, with  $V_r$  taking the role of  $V$ . Instead of computing  $V_{r+1}$  one element at a time, we can split the operation between  $P$  processors, each of which computes approximately  $1/P$  share of the total number of elements of  $V_{r+1}$  concurrently. The processors then send their part of the value function to a central processor which combines these into the whole value function, and sends it back to all processors before the beginning of the next iteration. If communication between processors is instantaneous, we can reduce computation time per

<sup>21</sup>The number of grid points is  $\prod_{k=1}^K n_k$ , where  $n_k$  is the number of grid points used in discretization of state space of state  $k$ .

<sup>22</sup>Solving dynamic programming problems in a distributed fashion is an old idea (see for example Bertsekas (1982)), and has been used in a variety of applications (Bertsekas (2012)). In economics parallel dynamic programming has been applied to solve large-scale problems (see for example Cai et al. (2015)), and a combination of software and hardware advances make this approach increasingly attractive.

iteration to  $1/P$  of the computation time on a single processor. In practice the performance gains are lower due to communication time and synchronisation delays.<sup>23</sup>

Parallel discrete VFI on a supercomputer proves successful in solving the models in this thesis. Unfortunately, there are scalability limits due to the curse of dimensionality. If our goal is to solve the models for more than two agents or to substantially increase the accuracy of the solution, the number of grid points becomes prohibitively great even if the number of available processors is unlimited. The communication time between processors (the time it takes to transfer the value function) becomes too great for parallelisation to be efficient, and we may even run out of computer memory. Every processor needs an updated copy of the whole value function at every iteration, and this is a potentially very large object.<sup>24</sup> In contrast, a parametric approximation of the value function can represent the value function much more compactly, which mitigates the problem of communication time. Furthermore, parametric approximations of the value function often provide more accurate solutions in a given computation time (Bertsekas (2012), Cai et al. (2013)).

The possibility of achieving a more accurate solution in a given time, and solving the model with more state variables due to better scalability is very attractive. For these reasons I implement a parametric value function algorithm.

### 3.5.3 Parametric approximation of the value function

The discrete VFI algorithm is modified by replacing the discrete approximation of the value function with a parametric approximation. The resulting algorithm is known as a fitted value iteration.

---

<sup>23</sup>The synchronisation delays stem from the fact that every processor needs to wait until all other processors finish the maximisation step in the VFI algorithm and the central processor combines the data, before beginning the next iteration. Parallelisation is particularly well suited for solving optimal policy models with heterogeneous agents, because the optimisation problem at every grid point is relatively complex. The program thus spends relatively more time solving the optimisation problems, compared to the time spent passing the results between processors.

<sup>24</sup>In the computation of the results presented in this thesis, the value function took up to 2 gigabytes of memory. The problem of communication time for a great number of grid points is particularly severe in the case of computational grids which were used by Cai et al. (2015). In this setting passing data between processors is much slower than in the case of using a supercomputer.



I use the standard choice of a linear approximation architecture. This means that the value function for each shock realisation is approximated as a linear combination of basis functions.

The choice of the basis functions is a crucial step. Products of orthogonal polynomials are commonly used in this setting (Cai and Judd (2014)). Complete Chebyshev polynomials were successfully applied in the solution of a multi-country growth model (Cai et al. (2013)). In the context of the model of Chapter 2, the closely related tensor products of Chebyshev polynomials (see Heer and Maussner (2009) or Judd (1998)) proved most appropriate in terms of providing a good approximation of value function after the first iteration of VFI, and in the following iterations.<sup>25</sup>

In general, convergence of the fitted VFI algorithm is difficult to achieve in practice.<sup>26</sup> Munos and Szepesvári (2008) present results on error-bounds of fitted value iteration after a finite number of iterations. These suggest that the algorithm can be expected to perform better with an increased richness of the family of basis functions, and with the number of points used for estimating the coefficients. Using a large number of points for estimating the coefficients is also intuitively appealing due to the relatively complex nature of the optimisation problem at every point in the state space, which may result in 'outliers' in the sense that the values found by numerical optimisation at some points are substantially worse than the true optimal values (given the current approximation of the value function).

Motivated by these observations, I implement a version of fitted VFI algorithm which allows the use of a great number of basis functions (up to several thousands), and the use of a large number of points in state space for estimating the coefficients in the linear approximation. For this purpose, standard linear algebra methods are often inadequate and I

<sup>25</sup>It is crucial to reach a good approximation in the first iteration (and in every consequent iteration). With the discount factor close to 1, the continuation value becomes quantitatively greater than the current return after several iterations. At every stage of VFI, we are then approximating a 'small term' plus a quantitatively greater polynomial of given form by a polynomial of the same form. Conventional measures of goodness of fit such as  $R^2$  often improve with subsequent iterations, even though the approximation of the true value function is in fact poor.

<sup>26</sup>The problem is that when the Bellman equation operator is combined with a projection of the value function into the linear space spanned by the basis functions, the resulting mapping is no longer a contraction (Bertsekas (2012)). Stachurski (2008) discusses non-expansive approximation schemes which can preserve the contraction property.

implement a learning algorithm (parallel gradient descent with a regularised least squares cost function) to overcome this problem.<sup>27</sup> The learning nature of the algorithm means that coefficients estimated in the previous iteration of VFI are updated, so that this information is used efficiently rather than discarded and the estimation becomes progressively faster if the VFI algorithm is stable. Other advantages include very efficient scalability with regard to parallelisation. The algorithm has potential applications in solving DSGE models with a relatively large number of state variables, or complex optimisation problems at each point in state space.

Unfortunately, despite experimenting with many different basis functions (high order tensor products of Chebyshev polynomials or complete Chebyshev polynomials, up to approximately 4,000 basis functions) and other parameters of the algorithm, the fitted VFI algorithm is not stable in practice. Modifications of the algorithm to achieve robust stability remain a topic for future research. A formal presentation and implementation details of both discrete and fitted VFI algorithms are given in the numerical appendix.

### 3.6 Conclusion

In this chapter I study the implications of heterogeneous productivity trends for optimal fiscal policy in a model with transfers. I explore the role of productivity trends in a two-period model with exogenous tax distortion, and compare the outcome to the model without transfers. In general, heterogeneity in levels becomes more important in determination of the optimal policy when transfers are introduced.

I numerically study optimal policy in the presence of heterogeneous productivity trends consistent with those observed in recent decades in the US and other advanced economies. These trends introduce a quantitatively significant trend component in the optimal policy, but usually less significant than in the model without transfers. In the horizon of several decades,

---

<sup>27</sup>This approach is often used in machine learning for estimating parameters of linear models when the number of observations and parameters is very high, and the data possibly contain errors (see for example Nguyen (2012)). The regularisation of the cost function is similar to the ridge regression used in econometrics, and it introduces a certain degree of robustness to errors in data.

the change in policy due to productivity trends quantitatively dominates the expected change in policy due to the effect of expenditure shocks.

One of the contributions of this chapter is developing a model which allows us to investigate the optimal policy for a range of possible scenarios regarding future labour productivity trends. It is well suited for making policy recommendations in periods in which a policy-maker predicts non-negligibly heterogeneous trends in earnings distribution. This may be particularly useful in the context of analysing the implications of technological change for the optimal policy.



# Chapter 4

## Open economy optimal fiscal policy<sup>1</sup>

### 4.1 Introduction

Following Lucas and Stokey (1983) and Aiyagari et al. (2002) (AMSS), many papers have studied optimal fiscal policy models where governments finance exogenous stochastic expenditures, and markets are either complete or incomplete. This literature has focused on the different implications of the structure of the markets for tax policy, as well as for debt dynamics and debt management.

A well known result of these models is that governments have an incentive to actively affect government bond prices in order to reduce the market value of outstanding debt. The Ramsey planner exploits the negative correlation between bond prices and deficits to achieve some degree of *fiscal insurance*, which allows the government to run looser budgets in the future to finance the current level of debt. This incentive is present both in complete and incomplete market environments.

Angeletos (2002) and Buera and Nicolini (2004) show that governments can effectively complete the markets with multiple bond maturities and they conclude that long bonds are more useful to achieve fiscal insurance because the price of long bonds can be affected more than short bonds by the Ramsey planner.

---

<sup>1</sup>This chapter is a joint work with Elisa Faraglia and Rigas Oikonomou.

The fiscal insurance channel is less explored in an open economy set up. Equiza-Goni et al. (2016) extend Angeletos (2002) to a two countries setting where the Ramsey planner cares equally about the two countries, and sets their individual fiscal policy and level of debt. In particular, they study how the correlation of the government spending shocks between the countries helps or prevents the government to attain the complete markets allocation using bonds of multiple maturities. In a simple three period model the authors show that the assumption of frictionless capital markets becomes a constraint for the Ramsey planner when the shocks are perfectly negatively correlated. The Ramsey planner would like to affect the bond prices in the two countries in opposite directions, but she is constrained by the no-arbitrage condition. The complete market outcome cannot be achieved and markets are effectively incomplete. The same authors extend their analysis (Equiza-Goni et al. (2017)) to a nominal setup and show that when shocks are perfectly negatively correlated across countries, then nominal bonds cannot complete the markets whereas indexed bonds can. The authors complement their theoretical analysis with some empirical evidence from 1999 to 2008 that shows that bond prices in a subset of countries in the Euro Area responded to aggregate shocks but not to idiosyncratic shocks. This evidence supports the theoretical findings that some degree of fiscal insurance in a frictionless capital market economy is achievable only when shocks are aggregate.

In view of these results, it is important to understand the implications of frictionless capital markets in an open economy for optimal policy, when we assume that markets are effectively incomplete. In particular, it is interesting to investigate how fiscal policy outcomes are affected by the cross country bond price equalisation constraint under different institutional settings.

In this article, we take an important step in this direction by studying optimal fiscal policy in an open economy setting with one-period bonds. More specifically, we extend the framework of AMSS into an open economy setting. We set up a two-country model of optimal policy, in which each country is a representative household production economy in the style of AMSS. The government of each country has to finance its stochastic expenditures by taxing labour, or issuing one-period risk-free bonds which are the only assets in the

economy. The asset purchases of each household are unrestricted in sign. Each government has to respect its own budget constraint, and the labour tax between the countries can vary. A Ramsey planner chooses a policy for all countries to maximise social welfare, and commits to this policy.

Our model encompasses three different institutional settings. The first setting is *autarky* (two separate closed economies). The second one is a *partial union* setting where financial markets are integrated. The households in each country can purchase bonds issued by either government. Third, we consider a *full union* model where transfers between governments are allowed in addition to international borrowing. The partial union setting is the main focus of this article, and we treat it as the baseline model in our presentation.

There are no frictions in the financial markets. The social planner faces a no-arbitrage constraint which is absent in the closed economy setting of AMSS. The returns on bonds of all countries must be equal in equilibrium. Combined with standard intertemporal optimality conditions of households in each country, this implies that the (expected) *growth* rate of marginal utility of consumption must be equal across all countries. Another interpretation is that the ratio of the expected next-period marginal utilities of any two countries equals the current ratio. This constraint implies that a difference between the marginal utility of consumption of households within any period is highly persistent. Therefore, consumption smoothing across countries is very important from the perspective of the Ramsey planner in an open economy. On the other hand, differences in labour supply between countries are not necessarily persistent.

The approach of this article is quantitative. We solve a standard Ramsey problem in a calibrated model in all three settings, and compare the optimal policy and welfare. Since the autarky case corresponds to the closed economy assumption, this approach allows us to quantitatively assess some of the consequences of moving to the open economy setting. We focus on the case of two ex ante identical countries. Shocks driven by a two-state Markov chain are either purely aggregate (positively correlated), or purely idiosyncratic (negatively

correlated).<sup>2</sup> These extreme shock processes allow us to demonstrate the differences between the three settings starkly. The Pareto weights are equal on both countries, which is motivated by the focus of this article being on the optimal policy response to expenditure shocks rather than redistribution. Our calibration targets long-run features of the US economy as in the previous two chapters.

If the expenditure shocks are purely aggregate, the optimal allocation and welfare are the same in all three settings. Therefore, we focus the discussion of our numerical results on the case of purely idiosyncratic shocks, keeping in mind that the results for autarky (in terms of allocation and welfare) also correspond to the other two settings if shocks are purely aggregate.

The results in the case of autarky are the same as in two separate closed economies in the model of AMSS. Taxes fluctuate considerably in response to shocks. For log preferences in consumption, the expected difference in consumption between the 'good' realisation (low expenditures) and the 'bad' realisation (high expenditures) of the shock process is 2.5% of steady-state consumption. The debt of the governments increases in periods of adverse shock realisation, and the allocation is persistently affected by the shocks. For the calibration used in this article, the changes in the stock of debt are quantitatively small.

In the case of the partial union international borrowing is possible. The fluctuations in the tax rate are not reduced. The change in government debt in response to shocks continues to be quantitatively small. The net external debt of the country with the high expenditure increases considerably, on average by 2.8% of GDP. The households in each country effectively borrow from each other in order to smooth consumption. The expected difference in consumption between the two shock realisations decreases from 2.5% of steady-state consumption to 0.11%.

One might intuitively expect that it would be better for governments to borrow from each other directly in order to smooth the distortionary taxes, rather than allow the tax rate to fluctuate and let households do the vast majority of the international borrowing to smooth

---

<sup>2</sup>These cases are the focus of the numerical examples given in Section 4.4. The model allows for the analysis of more general shock processes and ex ante heterogeneity in labour productivity, initial assets, and size of each country.



consumption. Our interpretation for why this is not the case is as follows. If a government borrows in bad times and saves in good times, this introduces a persistent asymmetry in the future spending needs of the governments when the correlation of shocks is negative. Taxes in the borrowing country would be persistently higher than in the other country. This is not optimal because of the convex distortionary cost of taxation. It is better to allow the tax rate to fluctuate in the present even at the cost of distortion in the short run, rather than to smooth taxes in the present and cause a greater and persistent expected distortion in the future. The fluctuations in labour supply increase relative to the closed economy, as the households in the country with lower expenditures (and taxes) work more, and lend resources to the households in the other country.

Both in the closed economy and in the partial union, the effect of shocks on the allocation and assets is persistent, and taxes and labour feature a random walk component. These results are standard in optimal fiscal policy models with incomplete markets.

In the case of the full union, only the aggregate government expenditures of all countries affect the set of feasible allocations. Because there is no aggregate uncertainty when shocks are purely idiosyncratic, consumption and taxes are perfectly smooth (across countries and time), which corresponds to the complete markets outcome. This case serves as a normative benchmark in our welfare comparisons.

In general, the welfare ranking of the three settings is not clear-cut, because the no-arbitrage constraint prevents the social planner from manipulating the interest rates in the two countries in the opposite directions. Because countries are *ex ante* identical and the initial asset holdings are zero, we expect the possible welfare gains from the absence of the no-arbitrage constraint to be small relative to the benefits from cross-country borrowing, which allows the households to achieve a smoother consumption profile.

We perform welfare analysis of the differences between the three settings for different degrees of risk-aversion of the households. While the differences in optimal policy between the three settings are significant, the welfare differences are minor. A move from the worst setting (autarky) to the best setting (full union) is associated with a utility gain equivalent to increasing consumption in autarky by 0.034%. The intermediate setting (partial union) is

better than autarky in terms of welfare, but the difference is quantitatively very small. These results are consistent with earlier results in the optimal fiscal policy literature. For example AMSS find small welfare differences between the settings of incomplete and complete markets in the closed economy, even though the differences in terms of policy are substantial.

The models in this article are very close to Equiza-Goni et al. (2016). The key differences are that here we assume effectively incomplete markets, only one-period bonds, and an infinite horizon. Moreover, this article is related to several branches of literature. On a technical level, the most closely related is the literature on optimal fiscal policy with heterogeneous agents and incomplete markets in a closed economy (Shin (2006), Bhandari et al. (2013)). These models also feature the constraint which equates the next-period ratio of expected marginal utility of different households to the current-period ratio of marginal utilities. The main differences are that in our article there are two governments with a separate budget constraint, and the tax rate is not constrained to be the same across households.

There are many papers in the new open economy macro literature which study optimal fiscal and monetary policy in a currency union. A non-exhaustive list of the major contributions includes Benigno (2004), Gali and Monacelli (2008), Ferrero (2009), and Farhi and Werning (2017). The closest to our article is Ferrero (2009), who also studies an optimal policy problem when taxes are distortionary and governments issue one-period debt. However, there are several important differences. We do not take the institutional arrangement as given, but instead compare the policy and welfare in autarky, partial union, and full union in an integrated setting, with an emphasis on the implications for the optimal policy of the no-arbitrage constraint in the open economy. Moreover, our solution approach is different. This literature relies on LQ approximations of the Ramsey problem to obtain a linear approximation of the policy function close to a deterministic steady state<sup>3</sup>. In contrast, we obtain a global approximation of the solution of the Ramsey problem by using numerical dynamic programming. This allows us to study optimal policy even in the long run, and in the face of substantial departures from a deterministic steady-state. Furthermore, we can compute accurate Monte Carlo expectations of social welfare in the different settings.

<sup>3</sup>Benigno and Woodford (2012) discuss the LQ approximation approach to optimal policy problems in detail.

A global solution method is particularly important in the context of the models studied in this article because of the magnitude and persistence of the impact of expenditure shocks on assets. In our numerical examples, in the partial union setting when shocks are purely idiosyncratic and initial assets are zero, the sample probability that the absolute value of net external debt of a country exceeds 100 percent of its GDP after 50 years is 23.7%.

Our model is also related to the literature on international risk-sharing and welfare cost of incomplete markets in economies without nominal rigidities. Representative examples include Cole and Obstfeld (1991), Van Wincoop (1994), Lewis (1996), Van Wincoop (1999), Lewis (2000), and Kim et al. (2003). The main differences are that these papers use endowment economies and do not solve an optimal policy problem. The cost of not being able to borrow internationally and the welfare cost of incomplete markets if international borrowing is possible, are usually quantitatively small unless shocks are highly persistent. This is consistent with the findings of our welfare analysis.

One of the contributions of this article is also computational. We solve a model of optimal fiscal policy in an open economy using a global solution method, which is parallel value function iteration on a discrete grid. The problem is similar to the problems studied in the previous two chapters, but there are several important differences. In particular, the related issues of non-convex feasible sets and local optima are more severe. We focus our discussion of the solution methods on overcoming these issues by use of exploration, and the choice of an appropriate interpolation scheme. We solve the model using Shepard's interpolation, and discuss the advantages relative to alternative interpolation schemes such as quadrilinear interpolation or B-splines in the numerical appendix.

The rest of this article is organised as follows. Section 4.2 contains the setup of the model, including the alternative specifications. In Section 4.3 we give a formulation of the Ramsey problem and briefly discuss the solution methods, focusing on the differences relative to the previous two chapters. Section 4.4 contains numerical examples. Section 4.5 concludes. Appendix C1 contains further details on the alternative specifications. Appendix C2 contains calibration details. Appendix C3 contains additional numerical results related to indeterminacy of the asset profile. Appendix C4 is the numerical appendix of this chapter.

## 4.2 A two-country model of fiscal policy

Our model extends the model of optimal fiscal policy with incomplete markets of Aiyagari et al. (2002) into a multi-country setting. We first describe the baseline setting of the partial union. Then we address the alternative settings of autarky and the full union.

There are two countries indexed by  $i = 1, 2$ , with a representative household in each country. The mass of country  $i$  households is  $\pi_i$ , with  $\pi_i > 0, \sum_i \pi_i = 1$ . This can be interpreted as the relative size of the country. Periods are indexed by  $t = 0, 1, \dots$ . Country  $i$ 's household preferences are represented by the time-separable utility function

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t^i, l_t^i), \quad (4.1)$$

where  $c_t^i$  and  $l_t^i$  are respectively consumption and labour supply of the household in country  $i$  in period  $t$ , and  $u$  is the instantaneous utility function.  $\mathbb{E}_0$  denotes the expectation conditional on time zero information and  $\beta \in (0, 1)$  is the discount factor. We focus on utility functions separable in consumption and labour  $u(c, l) = u(c) + u(l)$ , with  $u_c > 0, u_l < 0, u_{cc} < 0, u_{ll} < 0$ . Preferences of the social planner are represented by

$$\mathbb{E}_0 \sum_i \alpha_i \pi_i \sum_{t=0}^{\infty} \beta^t u(c_t^i, l_t^i), \quad (4.2)$$

where  $\alpha_i$  is a given Pareto weight on country  $i$  with  $\alpha_i \geq 0, \sum_i \alpha_i = 1$ .

The production function is CRS in labour supply. One unit of labour supplied by households in country  $i$  produces  $\theta_i > 0$  units of output. The maximum labour supply in each country is  $\bar{l} > 0$ . Labour is immobile between the countries.

Uncertainty in the economy is driven by a Markov chain with realisations  $s_t \in \{1, \dots, M\}$ , and  $M \times M$  transition matrix  $P$ . Per-capita government expenditure in country  $i$  is a function of this shock:  $g_t^i = g^i(s_t)$ . We use a superscript to denote a history of shock realisations  $s^t = (s_0, \dots, s_t)$ . The only source of uncertainty in the model are the government expenditure

shocks<sup>4</sup>. All of the following constraints hold in every period  $t = 0, 1, \dots$  and for every history of shock realisations.

The aggregate resource constraint (in all three settings) is

$$\sum_i \pi_i (c_t^i + g_t^i - \theta_i l_t^i) = 0. \quad (4.3)$$

Each country has its own distinct government, and a separate budget constraint. The policy of both governments is determined by a benevolent social planner, i.e. is perfectly coordinated in order to maximise the total social welfare in the partial union. We assume that the social planner (and both governments) are able to commit to a policy in period  $t = 0$ .

Both governments raise revenue through linear taxation of labour income at rate  $\tau_t^i$ . The tax rates can vary between the countries, but the social planner cannot directly transfer funds between the governments. The budget constraint of government  $i$  is

$$B_t^i = B_{t-1}^i R_{t-1}^i + \pi_i g_t^i - \pi_i \tau_t^i \theta_i l_t^i \quad (4.4)$$

where  $B_t^i$  is the amount of one-period risk-less bonds issued by country  $i$  in period  $t$ ,  $R_t^i$  is gross rate of return on the bonds and  $g_t^i$  are the per-capita government expenditures in country  $i$  in period  $t$ . As opposed to the previous two articles,  $B$  now stands for government debt rather than government assets.

The representative household in country  $i$  faces the budget constraint

$$\sum_j b_{i,t}^j = \sum_j b_{i,t-1}^j R_{t-1}^j + (1 - \tau_t^i) \theta_i l_t^i - c_t^i, \quad (4.5)$$

where  $b_{i,t}^j$  are the purchases of bonds issued by country  $j$  by households in country  $i$  in period  $t$ . The total amount of bonds issued by country  $j$  purchased by country  $i$  households is  $\pi_i b_{i,t}^j$ .

The government debt and asset purchases of the households are unrestricted in sign. Following Bhandari et al. (2013), we assume that the asset holdings of both the government

<sup>4</sup>It is straightforward to introduce other shocks such as TFP shocks or country-specific productivity shocks by making them a function of  $s_t$ , but for now we focus on expenditure shocks.

and the households are bounded, and we focus on interior solutions with respect to any borrowing constraints. The assumption of bounded assets rules out Ponzi schemes run by households and the government.

The market-clearing condition for bonds issued by country  $j$  is

$$\sum_i \pi_i b_{t,i}^j = B_t^j. \quad (4.6)$$

### 4.2.1 Competitive equilibrium and the Ramsey plan

In this section we give standard definitions related to the Ramsey plan, and we address the issue of indeterminacy of portfolio in a competitive equilibrium.

**Definition** (*allocation, asset profile, price system, tax policy*)

An *allocation* is a sequence  $\{c_t^i, l_t^i\}_{i,t}$  of consumption and labour supply in the two countries. An *asset profile* is a sequence of portfolio of households and government debt  $\{\{b_{j,t}^i\}_{i,j,t}, \{B_t^i\}_{i,t}\}$ . A *price system* is a sequence  $\{R_t^i\}_{i,t}$  of gross interest rates on the one-period bonds issued by each country. A *tax policy* is a sequence of tax rates on labour income in both countries  $\{\tau_t^i\}_t$ .

**Definition** (*competitive equilibrium*)

For a given initial asset profile  $\{\{b_{j,-1}^i\}_{i,j}, \{B_{-1}^i\}_i\}$ , and a given tax policy  $\{\tau_t^i\}_{i,t}$ , a *competitive equilibrium* is an allocation  $\{c_t^i, l_t^i\}_{i,t}$ , an asset profile  $\{\{b_{j,t}^i\}_{i,j,t}, \{B_t^i\}_{i,t}\}$  and a price system  $\{R_t^i\}_{i,t}$ , such that:

1. The allocation  $\{c_t^i, l_t^i\}_{i,t}$  and portfolio of the households  $\{b_{j,t}^i\}_{i,j,t}$ , maximise the utility (4.1) of household in each country, subject to the respective budget constraint (4.5).
2. The resource constraint (4.3) is satisfied.
3. The governments' budget constraints (4.4) are satisfied.
4. Market-clearing in the asset market (4.6) is satisfied.

5. The asset profile is bounded.

A *feasible tax policy* is a tax policy for which a competitive equilibrium exists (this implies the satisfaction of all constraints). The market-clearing condition holds in period  $t = -1$ , i.e. the initial portfolio of households is consistent with the initial level of government debt in both countries.

The Ramsey plan (socially optimal competitive equilibrium) is a pair of a tax policy and a competitive equilibrium consistent with the tax policy, which maximise the social planner's objective function (over all feasible tax policies). A formal definition follows.

**Definition** (*socially optimal competitive equilibrium - Ramsey plan*)

Given an initial asset profile  $\{\{b_{j,-1}^i\}_{i,j}, \{B_{-1}^i\}_i\}$ , a *socially optimal competitive equilibrium* (Ramsey plan) is a tax policy  $\{\tau_t^{i*}\}_{i,t}$ , an allocation  $\{c_t^{i*}, l_t^{i*}\}_{i,t}$ , an asset profile  $\{\{b_{j,t}^{i*}\}_{i,j,t}, \{B_t^{i*}\}_{i,t}\}$  and a price system  $\{R_t^{i*}\}_t$  such that:

1. Given the initial asset profile  $\{\{b_{j,-1}^i\}_{i,j}, \{B_{-1}^i\}_i\}$  and the tax policy  $\{\tau_t^{i*}\}_{i,t}$ , the other elements of the socially optimal competitive equilibrium constitute a competitive equilibrium.
2. Given the initial asset profile  $\{\{b_{j,-1}^i\}_{i,j}, \{B_{-1}^i\}_i\}$ , there is no other feasible tax policy  $\{\tau_t^i\}_t \neq \{\tau_t^{i*}\}_t$ , for which a competitive equilibrium exists with a strictly greater social welfare.

A competitive equilibrium satisfies several properties which are useful for characterising the optimal policy (Ramsey plan) and solving the model.

In any interior competitive equilibrium, the following no arbitrage condition imposing equality of interest rates on both countries' bonds must hold in every period.<sup>5</sup>

$$R_t^1 = R_t^2. \quad (4.7)$$

<sup>5</sup>This follows from the assumption that the assets are risk-less and identical in terms of purchase price. By contradiction, suppose that we are in an interior competitive equilibrium, and the condition does not hold, i.e. without loss of generality  $R_t^1 > R_t^2$  in some period  $t$ . Then every household  $i$  would be able to increase its utility by decreasing the purchases of the asset with the lower rate of return ( $b_{i,t}^1$ ) and increasing purchases of the asset with the higher rate of return ( $b_{i,t}^2$ ).

In a competitive equilibrium, households in each country  $i$  maximise their expected utility (4.1), subject to their budget constraint (4.5). The standard<sup>6</sup> intratemporal and intertemporal first-order conditions of the households' problems are

$$\theta_i(1 - \tau_t^i) = -\frac{u_{l,t}^i}{u_{c,t}^i} \quad (4.8)$$

and

$$R_t^i = \frac{u_{c,t}^i}{\beta \mathbb{E}_t(u_{c,t+1}^i)}. \quad (4.9)$$

In a competitive equilibrium, the asset profile is indeterminate. More specifically, the gross external debt of each country is not determined, and only the *net* external debt of each country is determined. We define the *net* external debt of country 1 as  $b_{1,t}^{ex} \equiv \pi_2 b_{t,2}^1 - \pi_1 b_{t,1}^2$ , with external debt of country 2 being  $b_{2,t}^{ex} \equiv -b_{1,t}^{ex}$ . We show that starting in a competitive equilibrium, as long as the *net* external debt is fixed, we can find an infinite number of new asset profiles consistent with the original equilibrium allocation (**Result 1**). This result relies on the assumptions that the bonds issued by the two countries are perfect substitutes because they offer the same return due to the no-arbitrate condition. The result is similar in spirit to Bhandari et al. (2013) who show an indeterminacy of asset holdings in a one-country model with heterogeneous agents and uniform transfers.

**Result 1** (*Indeterminacy of foreign debt level*)

Assume that an allocation  $\{c_t^{i*}, l_t^{i*}\}_{i,t}$ , an asset profile (A)  $\{\{b_{j,t}^{i*}\}_{i,j,t}, \{B_t^{i*}\}_{i,t}\}$  and a price system  $\{R_t^{i*}\}_t$  are a competitive equilibrium given a tax policy  $\{\tau_t^i\}_i$ . Take any  $k \in \mathbb{R}$  and construct a new asset profile (B)  $\{\{\hat{b}_{j,t}^i\}_{i,j,t}, \{\hat{B}_t^i\}_{i,t}\}$  such that

- (i)  $\{\{\hat{b}_{j,t}^i\}_{i,j,t}, \{\hat{B}_t^i\}_{i,t}\} = \{\{b_{j,t}^{i*}\}_{i,j,t}, \{B_t^{i*}\}_{i,t}\}$  in all periods  $t \neq s$
- (ii)  $\hat{b}_{2,s}^1 = b_{2,s}^{1*} + k, \hat{b}_{1,s}^2 = b_{1,s}^{2*} + \frac{\pi_2}{\pi_1} k$
- (iii)  $\hat{b}_{1,s}^1 = b_{1,s}^{1*} - \frac{\pi_2}{\pi_1} k, \hat{b}_{2,s}^2 = b_{2,s}^{2*} - k$

---

<sup>6</sup>Appendix A1 contains derivation of the first-order conditions of the problem of households in a similar setting.



$$(iv) \hat{B}_s^1 = B_s^{1*}, \hat{B}_s^2 = B_s^{2*}$$

The original equilibrium allocation  $\{c_t^{i*}, l_t^{i*}\}_{i,t}$ , the new asset profile  $\{\{\hat{b}_{j,t}^i\}_{i,j,t}, \{\hat{B}_t^i\}_{i,t}\}$ , and the original price system  $\{R_t^{i*}\}_t$  are a competitive equilibrium given the tax policy  $\{\tau_t^i\}_i$ .

*Proof.* The original equilibrium allocation combined with the new asset profile (B) satisfies the budget constraints of both countries' households. To see this, note that all constraints in periods except for  $s, s+1$  are unchanged. In period  $s$  the constraint in country 1 is reduced to the original constraint as well<sup>7</sup>

$$\begin{aligned} \hat{b}_{1,s}^1 + \hat{b}_{1,s}^2 &= (\hat{b}_{1,s-1}^1 + \hat{b}_{1,s-1}^2)R_{s-1}^* + (1 - \tau_s^1)\theta_1 l_s^{1*} - c_s^{1*} \\ b_{1,s}^{1*} - \frac{\pi_2}{\pi_1}k + b_{1,s}^{2*} + \frac{\pi_2}{\pi_1}k &= (b_{1,s-1}^{1*} + b_{1,s-1}^{2*})R_{s-1}^* + (1 - \tau_s^1)\theta_1 l_s^{1*} - c_s^{1*} \\ b_{1,s}^{1*} + b_{1,s}^{2*} &= (b_{1,s-1}^{1*} + b_{1,s-1}^{2*})R_{s-1}^* + (1 - \tau_s^1)\theta_1 l_s^{1*} - c_s^{1*}. \end{aligned}$$

In period  $s+1$  the constraint is again reduced to the original one

$$\begin{aligned} \hat{b}_{1,s+1}^1 + \hat{b}_{1,s+1}^2 &= (\hat{b}_{1,s}^1 + \hat{b}_{1,s}^2)R_s^* + (1 - \tau_{s+1}^1)\theta_1 l_{s+1}^{1*} - c_{s+1}^{1*} \\ b_{1,s+1}^{1*} + b_{1,s+1}^{2*} &= (b_{1,s}^{1*} - \frac{\pi_2}{\pi_1}k + b_{1,s}^{2*} + \frac{\pi_2}{\pi_1}k)R_s^* + (1 - \tau_{s+1}^1)\theta_1 l_{s+1}^{1*} - c_{s+1}^{1*} \\ b_{1,s+1}^{1*} + b_{1,s+1}^{2*} &= (b_{1,s}^{1*} + b_{1,s}^{2*})R_s^* + (1 - \tau_{s+1}^1)\theta_1 l_{s+1}^{1*} - c_{s+1}^{1*} \end{aligned}$$

Similarly, we can show that country 2 household can afford the original allocation if it chooses the new asset profile. In the same way we can show that any allocation which is feasible under the new asset profile (B) is also feasible under the old asset profile (A), when the original equilibrium allocation  $\{c_t^{i*}, l_t^{i*}\}_{i,t}$  was chosen. In summary, the original allocation with the new asset profile (B) is both feasible and optimal, and it thus continues to solve the utility maximisation problem of households in both countries.

We now turn to the remaining conditions for a competitive equilibrium. The resource constraint is still satisfied because the allocation is the same. The budget constraint of the government is satisfied because the government debt was not changed. The market-clearing

<sup>7</sup>We use the no-arbitrage condition to simplify the expression and define  $R_t^* \equiv R_t^1 = R_t^2$ .

condition for bonds of both countries is still satisfied. For country 1, the market-clearing condition in period  $s$  is

$$\begin{aligned} B_s^{1*} &= \hat{B}_s^1 = \pi_1 \hat{b}_{1,s}^1 + \pi_2 \hat{b}_{2,s}^1 \\ &= \pi_1 (b_{1,s}^{1*} - \frac{\pi_2}{\pi_1} k) + \pi_2 (b_{2,s}^{1*} + k) \\ &= \pi_1 b_{1,s}^{1*} + \pi_2 b_{2,s}^{1*} \end{aligned}$$

which is the original market-clearing condition. The same is true for country 2. Finally, since markets clear at the original price system, it continues to be an equilibrium price system.  $\square$

As opposed to the asset levels, the net external debt is determined in equilibrium.<sup>8</sup> We treat the net external debt of country 1  $b_1^{ex}$  as a state variable. Without loss of generality (Result 1) we can use the normalisation  $b_{1,t}^2 \equiv 0$  to recover the asset profile, or any other convenient normalisation. Since the asset profile is indeterminate in equilibrium, our prime interest is in the evolution of the real allocation, the tax rate, the debt of the governments and the net external debt. Conditional on a normalisation, the changes in the asset profile are also meaningful to analyse. To illustrate more clearly the indeterminacy of the asset profile (portfolio of households), and how different asset profiles can support the same real allocation, Appendix C2 shows optimal portfolios in the partial union for two different normalisations, for a numerical example with purely idiosyncratic shocks of Section 4.4.

The fact that we can adopt a normalisation in which one of the cross-country bonds purchases is zero allows us to express the Ramsey problem with three assets only. In a recursive formulation of the problem we need to keep track of only three assets rather than the whole portfolio of households<sup>9</sup>.

<sup>8</sup>This is true to the extent that the real allocation is uniquely determined in equilibrium. Any change in the asset profile which does not keep the net external debt fixed (as the changes considered in Result 1 did) implies a failure of the market-clearing in the asset market at the going interest rate. This implies a change of interest rate, which is tied to the real allocation in equilibrium through households' Euler equations.

<sup>9</sup>This is similar to the second article in this thesis (Chapter 3), where the indeterminacy of asset levels also allowed a reduction of the dimensionality of the state space by one.

Since the net external debt is of key importance in this model, it is useful to derive its law of motion and use it as a state variable. For country  $i$  this is<sup>10</sup>

$$b_{i,t}^{ex} = b_{i,t-1}^{ex} R_{t-1}^i + \pi_i [g_t^i + c_t^i - \theta_i l_t^i]. \quad (4.10)$$

We now discuss the two closely related alternative model specifications which serve for comparisons both in terms of policy and welfare. The first case (full union) extends the model by adding transfers between governments. The second case (autarky) is the closed economy case which rules out international borrowing.

### 4.2.2 Full union

We model the full union in the same vein as the baseline partial union model, which we have described earlier. The only substantial difference is that the social planner can use transfers between the two governments in addition to the already available policy instruments. We define  $T_{j,t}^i$  to be the transfer going from government  $i$  to government  $j$  in period  $t$ , which is unrestricted in sign.<sup>11</sup> The budget constraints of the governments are modified by the addition of transfers

$$B_t^i = B_{t-1}^i R_{t-1}^i + \pi_i g_t^i - \pi_i \tau_t^i \theta_i l_t^i + \sum_j (T_{i,t}^j - T_{j,t}^i) \quad (4.11)$$

Summing this over all countries  $i$ , we obtain a single pooled budget constraint for the whole union

$$\sum_i B_t^i = \sum_i B_{t-1}^i R_{t-1}^i + \sum_i \pi_i g_t^i - \sum_i \pi_i \tau_t^i \theta_i l_t^i + \sum_i \sum_j (T_{i,t}^j - T_{j,t}^i) \quad (4.12)$$

<sup>10</sup>To show this, subtract the household budget constraint of country  $i$  from the budget constraint of country  $i$  government. Then use the market-clearing condition and the no-arbitrage condition to simplify the resulting expression.

<sup>11</sup>With the full set of transfers unrestricted in sign there is indeterminacy in the sense that infinitely many choices of transfers correspond to the same redistribution of resources between the governments. For example, increasing  $T_{2,1}$  and  $T_{1,2}$  by the same amount does not affect the resources available to either government. We can assume for example  $T_{1,t}^2 \equiv 0$  to pin down the level of transfers, but we do not adopt any such normalisation because we are primarily interested in the allocation and not in the transfers that are used to implement it.

For any asset profile which satisfies the pooled constraint (4.12), we can find a profile of transfers  $\{T_j^i\}_{i,j}$  such that the individual governments' budget constraints are satisfied. For this reason, we can replace the individual budget constraints with the pooled budget constraint when solving the Ramsey problem. After we have found an optimal allocation we can recover a profile of transfers such that the individual budget constraints of all governments are satisfied.

We can simplify the pooled constraint further. The transfer payments cancel out since the summation range is the same for both  $i$  and  $j$ . By the no arbitrage condition, the interest rate on all bonds is the same in any competitive equilibrium:  $R_t^i = R_t$ . Finally, since the debt of the individual governments only enters the pooled constraint as the sum of debt of all countries (after we use the no-arbitrage condition), we can replace the sum of debt with a single government debt of the union  $B_t \equiv \sum_i B_t^i$ . The pooled government budget constraint in terms of the single asset is

$$B_t = B_{t-1}R_{t-1} + \sum_i \pi_i g_t^i - \sum_i \pi_i \tau_t^i \theta_i l_t^i. \quad (4.13)$$

This constraint reflects the fact that, from the perspective of the government's budget constraints, the distribution of government debt between countries does not matter if resources are pooled by the use of transfers between governments.

Because the bonds issued by any government are perfect substitutes in this model, we can formulate the country  $i$  representative household's problem as choosing the total amount of bonds purchased  $b_i \equiv \sum_j b_i^j$ , rather than keeping track of the amount of bonds purchased from each government.<sup>12</sup> The budget constraint of the representative household in country  $i$  thus becomes

$$b_i = b_{i,t-1}R_{t-1} + (1 - \tau_t^i)\theta_i l_t^i - c_t^i. \quad (4.14)$$

Since the distribution of government debt between individual governments matters neither from the perspective of the (budget constraint of) households nor from the perspective of the

<sup>12</sup>In the baseline setting of partial union, households are (in equilibrium) also indifferent between buying bonds issued by different governments. However, we still need to keep track of their portfolios because these matter from the perspective of the separate budget constraints of the governments.

governments, we can formulate the social planner's problem using a single union bond  $B_t$ . The market-clearing condition for this bond is

$$\sum_i \pi_i b_{t,i} = B_t. \quad (4.15)$$

The formal definitions of a competitive equilibrium and Ramsey plan, and a sequential and recursive version of the Ramsey problem are straightforward modifications of those in the partial union setting, and are relegated to Appendix C1. We proceed with a brief discussion of the implications of adding transfers into the model.

The baseline partial union model can be viewed as a special case of the more general model with transfers, with the additional constraint  $T_{j,t}^i = 0$  for all  $i, j, t$ . Any competitive equilibrium in the partial union, and the socially optimal equilibrium in particular, can also be attained in the full union (if all transfers are set to zero). It follows that the expected social welfare in the full union is at least as great as in the partial union.<sup>13</sup>

Intuitively, if shocks are not perfectly correlated there is a possibility for risk-sharing, and the expected social welfare with transfers should be strictly greater than without transfers as this additional instrument allows the government to complete the markets. In the special case of purely idiosyncratic shocks which we focus on in the numerical examples, the aggregate expenditure of the union is constant, and consumption is perfectly smoothed.

### 4.2.3 Autarky

In autarky, cross-country borrowing is ruled out, so that  $b_j^i = 0$  for  $i \neq j$ . Since there is no link between the countries, the optimal policy for both countries can be found by solving a one-country optimal policy model for each country in isolation. The one-country problem is essentially the same as in the closed economy model of AMSS, and we therefore relegate a

<sup>13</sup>Because the policy of the governments is perfectly coordinated (chosen by the social planner) to maximise welfare at the level of the whole union, there is no possible welfare gain from the presence of the constraints. If the policy were not coordinated, the constraints on transfers could be welfare-improving due to restricting the actions of the individual governments.

more formal treatment of the model to Appendix C1. Table 4.1 summarises the differences between the three versions of the model.

Table 4.1 Differences between the models

	<b>Integrated asset market</b>	<b>International transfers</b>
<b>Autarky</b>	no ( $b_j^i = 0$ for $i \neq j$ )	no ( $T_j^i = 0$ for $i \neq j$ )
<b>Partial union</b>	yes	no ( $T_j^i = 0$ for $i \neq j$ )
<b>Full union</b>	yes	yes

The key difference between autarky and the (partial and full) union is that in autarky, the no arbitrage constraint which equates the returns on the bonds issued by the two governments is absent.

#### 4.2.4 The implications of the no-arbitrage constraint

In the union setting (both partial and full), the no-arbitrage condition together with the intertemporal optimality conditions of the households imply that the *expected* growth rate (not level) of marginal utility has to be levelled across all countries

$$\frac{\mathbb{E}_t(u_{c,t+1}^1)}{u_{c,t}^1(s^t)} = \frac{\mathbb{E}_t(u_{c,t+1}^2)}{u_{c,t}^2(s^t)}. \quad (4.16)$$

To see the implications of this constraint more clearly, it is useful to rewrite it as

$$\rho(s^t) \equiv \frac{u_{c,t}^1(s^t)}{u_{c,t}^2(s^t)} = \frac{\mathbb{E}_t(u_{c,t+1}^1)}{\mathbb{E}_t(u_{c,t+1}^2)}. \quad (4.17)$$

In any period, the ratio of the expected next-period marginal utilities of consumption of the two countries equals the current ratio. This constraint implies that any deviations from consumption smoothing between the countries are persistent.

Consider that in period  $t$ , the ratio of marginal utilities changes from  $\rho_{t-1} = 1$  to  $\rho_t = 1 + k$ , with  $k > 0$  in response to an asymmetric expenditure shock. A consumption plan for histories  $s^{t+1}$  which keeps the ratio of marginal utilities the same as in period  $t$  for all shock realisations satisfies the constraint, so it is possible to guarantee not increasing the deviation from the

original ratio of marginal utilities further. However, a consumption plan which brings the ratio back to 1 in period  $t + 1$  for all shock realisations violates the constraint. The only way to return to the original ratio is to increase the deviation from the original ratio even more for some histories  $s^{t+1}$ .

In contrast, differences in labour supply are not propagated into future periods through this channel. The asymmetry between the persistence of differences in labour supply and in consumption is reflected in the optimal policy in the partial union, in which consumption is considerably smoothed between countries, but labour supply is not. We explore this further in the discussion of our numerical results.

The fact that the no-arbitrage constraint is absent in autarky implies that the welfare ranking of the three settings is not clear. In particular, autarky may not be the worst in terms of social welfare. The households and the governments cannot borrow internationally, which prevents standard risk-sharing benefits. However, the government can affect bond returns to achieve some fiscal insurance due to the absence of the no arbitrage constraint. There are equilibrium allocations that can be implemented in autarky but not when international borrowing is allowed.

Intuitively, the welfare cost of not being able to achieve a different return on bonds issued by the two countries may be significant particularly if the countries are ex ante heterogeneous, and their asset holdings are different.

## 4.3 The Ramsey problem

### 4.3.1 The Ramsey problem (sequential version)

In this subsection we present a formulation of the Ramsey problem in the space of sequences for the partial union model. The Ramsey problems for the alternative settings (autarky, full union) are relegated to Appendix C1.

Following standard practice we use the primal approach, i.e. we use the households' first-order conditions to eliminate the interest rate and tax policy in all constraints, and we solve for the allocation directly. The tax rates in the two countries and the interest rate can

be obtained from the allocation using the first-order conditions of the households. Using Walras' law we can ignore one constraint, in this case the law of motion for external debt of country 2 (which represents the budget constraint of households in that country). The optimal policy is found by maximising the social planner's objective function (4.2) subject to (4.4), (4.3), (4.7), (4.8), (4.9) and country 1 version of constraint (4.10). Putting this all together, we obtain the following sequential formulation of the Ramsey problem:

$$\max_{\{c_t^i(s^t), l_t^i(s^t), B_t^i(s^t), b_{1,t}^{ex}(s^t)\}_{t=0, \forall i, s^t}}^{\infty} \mathbb{E}_0 \sum_i \alpha_i \pi_i \sum_{t=0}^{\infty} \beta^t u(c_t^i, l_t^i)$$

subject to

$$B_t^i(s^t) = B_{t-1}^i(s^{t-1}) \frac{u_{c,t-1}^i(s^{t-1})}{\beta \mathbb{E}_{t-1}(u_{c,t}^i)} + \pi_i g_t^i(s^t) - \pi_i \left( 1 + \frac{u_{l,t}^i(s^t)}{\theta_i u_{c,t}^i(s^t)} \right) \theta_i l_t^i(s^t), \forall i, t, s^t \quad (4.18)$$

$$0 = \sum_i \pi_i (c_t^i(s^t) + g_t^i(s^t) - \theta_i l_t^i(s^t)), \forall t, s^t \quad (4.19)$$

$$b_{1,t}^{ex}(s^t) = b_{1,t-1}^{ex}(s^{t-1}) \frac{u_{c,t-1}^1(s^{t-1})}{\beta \mathbb{E}_{t-1}(u_{c,t}^1)} + \pi_1 [g_t^1(s^t) + c_t^1(s^t) - \theta_1 l_t^1(s^t)], \forall t, s^t \quad (4.20)$$

$$\frac{u_{c,t}^1(s^t)}{\mathbb{E}_t(u_{c,t+1}^1)} = \frac{u_{c,t}^2(s^t)}{\mathbb{E}_t(u_{c,t+1}^2)}, \forall t, s^t \quad (4.21)$$

### 4.3.2 The Ramsey problem (recursive formulation)

As in the previous two chapters, we follow the approach of Kydland and Prescott (1980). The value function for periods  $t \geq 1$  captures the value at the start of the period before current-period shock is realised, and we find a state-contingent plan for the current period. The notation is standard in the sense that for any variable  $x$ ,  $x'$  denotes the next-period value and  $x_-$  the last-period value.

The state variables are:

- $\mathbf{b} = (B^1, B^2, b_1^{ex})$ : last-period assets (government debt country 1, government debt of country 2, and net foreign debt of country 1)



- $\mathbf{u} = (u_c^1, u_c^2)$ : last-period marginal utility of consumption of households in both countries
- $s_-$ : last-period shock realisation

The choice variables are  $[c^i(s), l^i(s), B^i(s)']$  for every possible realisation of shock  $s$  and for both countries  $i$ , and  $b_1^{ex}(s)'$  for country 1. The summation  $\sum_s$  means summation over all possible shock realisations. Notation  $\mathbf{b}(s)' \equiv (B^1(s)', B^2(s)', b_1^{ex}(s)')$  for the vector of current-period asset purchases and  $\mathbf{u}^c(s) \equiv (u_c^1[c^1(s), l^1(s)], u_c^2[c^2(s), l^2(s)])$  for the vector of current-period marginal utilities of consumption (both vectors depend on the realised shock) is used. The notation  $u^i(s) \equiv u(c^i(s), l^i(s))$  is adopted, and we use the same notation for marginal utilities.

The Bellman equation for periods  $t \geq 1$  is

$$V(\mathbf{b}, \mathbf{u}, s_-) = \max_{[c^i(s), l^i(s), B^i(s)']_{i,s}, [b_1^{ex}(s)']_s} \sum_s Pr(s|s_-) \left[ \sum_{i=1}^2 \alpha_i \pi_i u(c^i(s), l^i(s)) + \beta V(\mathbf{b}(s)', \mathbf{u}^c(s), s) \right]$$

subject to

$$B^i(s)' = B^i \frac{u_c^i}{\beta \sum_s Pr(s|s_-) u_c^i(s)} + \pi_i g^i(s) - \pi_i \left( 1 + \frac{u_l^i(s)}{\theta_i u_c^i(s)} \right) \theta_i l^i(s), \forall i, s \quad (4.22)$$

$$0 = \sum_i \pi_i (c^i(s) + g^i(s) - \theta_i l^i(s)), \forall s \quad (4.23)$$

$$b_1^{ex}(s)' = b_1^{ex} \frac{u_c^1}{\beta \sum_s Pr(s|s_-) u_c^1(s)} + \pi_1 [g^1(s) + c^1(s) - \theta_1 l^1(s)], \forall s \quad (4.24)$$

$$\frac{u_c^1}{\sum_s Pr(s|s_-) u_c^1(s)} = \frac{u_c^2}{\sum_s Pr(s|s_-) u_c^2(s)} \quad (4.25)$$

The period 0 problem involves choosing an allocation and assets such that the period 0 social welfare plus the continuation value (which is given by the approximate value function) is maximised:

$$\max_{[c_0^i, l_0^i, B_0^i]_{i,s}, [b_{1,0}^{ex}]_s} \sum_{i=1}^2 \alpha_i \pi_i u(c^i(s), l^i(s)) + \beta V(B_0^1, B_0^2, b_{1,0}^{ex}, u_{c,0}^1, u_{c,0}^2, s_0)$$

subject to

$$B_0^i = B_{-1}^i + \pi_i g_0^i - \pi_i \left( 1 - \frac{u_{l,0}^i}{\theta_i u_{c,0}^i} \right) \theta_i l_0^i, \forall i \quad (4.26)$$

$$0 = \sum_i \pi_i (c_0^i + g_0^i - \theta_i l_0^i) \quad (4.27)$$

$$b_{1,0}^{ex} = b_{1,-1}^{ex} + \pi_1 [g_0^1 + c_0^1 - \theta_1 l_0^1] \quad (4.28)$$

We assume that  $t = -1$  period gross rate of return on the initial assets  $(B_{-1}^1, B_{-1}^2, b_{1,-1}^{ex})$  equals one. Note that there is no uncertainty in this problem because period 0 shock  $s_0$  is known at the time of deciding period 0 allocation and asset purchases, and the value function captures the expected return at the beginning of period, before the shock realisation is observed.

### 4.3.3 Solution methods

We discuss the solution of the partial union setting. As in the previous two chapters, our solution approach is based on using parallel value function iteration on a discrete grid, using a recursive formulation of the social planner's problem with reduced dimensionality (given in Appendix C4). The basic structure of the algorithm is similar to the ones which were used in the first two chapters of this thesis. We focus here on the key differences.

Firstly, in the current model there is no productivity growth, which makes it comparatively easier to solve. There is one fewer state variable, and we do not need to take grid boundaries as wide as in the presence of trends because there is less variation in states over time (so we can afford denser grids given a number of grid points).

Secondly, there are now three assets (as opposed to two) which we have to keep track of as state variables, which makes the problem more difficult to solve. This additional state is continuous, unlike the discrete 'time' variable in the previous model.

Lastly, the tax rate can vary between countries. This implies one fewer constraint and an increase in the number of free variables by one per shock realisation. This makes the model

more challenging to solve, in particular with respect to finding an initial guess, and dealing with non-convexities.

In Appendix C4 we discuss the most important features of the solution algorithm which are substantially different from the first two chapters. In particular, we focus on dealing with the consequences of non-convexities (and the related problem of local optima) which are more severe in this chapter than in the previous two. In this context we discuss exploration and the performance of various interpolation schemes such as quadrilinear interpolation, B-splines, and Shepard's interpolation, which is the best-performing method found for solving this model.<sup>14</sup>

## 4.4 Numerical examples

In this section, we present our numerical results. We solve a calibrated version of the model in all three settings: autarky, partial union, and full union. Comparing the outcomes in the three settings allows us to quantitatively assess the implications of the open economy assumption, as the autarky setting corresponds to two closed economies. Firstly, we focus on policy comparisons. Secondly, we analyse the welfare implications of the three settings.

As is common in the optimal fiscal policy literature we use CRRA preferences. The instantaneous utility of household in country  $i$  is

$$u(c^i, l^i) = \frac{c^{i(1-\sigma)}}{1-\sigma} - B \frac{l^{i(1+\gamma)}}{1+\gamma}, \quad (4.29)$$

where  $B > 0$ ,  $\sigma > 0$ ,  $\gamma > 0$ , and  $\sigma = 1$  corresponds to logarithmic utility of consumption.

---

<sup>14</sup>The numerical results obtained with one of the other two interpolation methods are very similar in terms of sample averages, which we focus our discussion on. However, the series for tax rates usually display a more erratic behaviour, reflecting the fact that even a small inaccuracy in terms of allocation can result in a relatively large inaccuracy in terms of the tax rate.

### 4.4.1 Calibration

In the examples of this section we focus on the case of ex ante identical countries (equal size and productivity). We also assume equal Pareto weights.

The majority of parameters are the same as in the calibrations of the models used in the previous two chapters, and thus target features of the US economy in recent decades. Appendix C3 contains a table which summarises all calibration targets and parameters. Here we only discuss the elements of the calibration which are most important for the optimal policy response to shocks and the welfare analysis: the government expenditure process, and preferences.

**Government expenditures** The central government expenditure is  $\bar{g}$ . In the examples of this section the government expenditures in each country take on two values; low ( $g_L$ ) or high ( $g_H$ ), with  $g_L < \bar{g} < g_H$ . Shocks are either purely idiosyncratic, or purely aggregate. In the case of purely idiosyncratic shocks, whenever the expenditure is low in one country it is high in the other, so the aggregate expenditure is constant. In the case of purely aggregate shocks, the expenditure is always either low or high in both countries.

An underlying Markov chain with realisations  $s \in \{1, 2\}$  and a transition matrix  $P$  determines which of the two possible shock realisations occurs in each country. Table 4.2 summarises the two cases.

Table 4.2 Calibration of shocks

Markov chain realisation	Country 1	Country 2
Purely aggregate shocks (correlation = 1)		
$s_t = 1$	$g_t^1 = g_L$	$g_t^2 = g_L$
$s_t = 2$	$g_t^1 = g_H$	$g_t^2 = g_H$
Purely idiosyncratic shocks (correlation = -1)		
$s_t = 1$	$g_t^1 = g_L$	$g_t^2 = g_H$
$s_t = 2$	$g_t^1 = g_H$	$g_t^2 = g_L$

Following the first two chapters, we set the level of government expenditures such that the labour tax rate is 25 percent in a steady-state with zero assets and mean government

expenditure  $\bar{g} = \frac{g_L + g_H}{2}$  in each country.<sup>15</sup> We choose  $g_L = 0.93\bar{g}$  and  $g_H = 1.07\bar{g}$ , following Faraglia et al. (2008). We assume that shocks are i.i.d., so each row of the transition matrix is  $(0.5, 0.5)$ .

There are two reasons for our focus on the two extreme cases of correlation of expenditures. Firstly, the optimal policy for these shocks processes is markedly different between the three settings, and it allows us to underscore the differences. Secondly, the case of purely idiosyncratic shocks is particularly interesting with regard to welfare analysis. It offers the maximum risk-sharing potential, in the sense that the variance of the aggregate expenditures in the union is zero, and the social planner can achieve a perfectly smooth consumption in the full union.

**Preferences** The extent to which households are averse to fluctuations in consumption is important both with respect to policy and welfare. We solve the model for commonly used utility logarithmic in consumption ( $\sigma = 1$ ), and also for a higher degree of risk-aversion ( $\sigma = 2, \sigma = 3$ ).

#### 4.4.2 Results

We compare the three settings when shocks are purely idiosyncratic. Because countries are ex ante identical and the Pareto weights are equal, if shocks are purely aggregate then the allocation and welfare in both countries are the same in all three settings.<sup>16</sup> For this reason, we focus the following discussion on the case of purely idiosyncratic shocks, bearing in mind that the allocation and welfare in autarky also correspond to all three settings when shocks are purely aggregate.

<sup>15</sup>This level is consistent with long-run US tax rate as discussed in Chapter 2. A similar value is used for example by Faraglia et al. (2008).

<sup>16</sup>We verified this claim numerically for the settings of partial union and autarky, in which case the optimal international borrowing was zero. The intuition is that if there are two ex ante identical countries which face the same shock realisations, no increase in social welfare can be achieved by allowing them to borrow from each other or making international transfers (because of decreasing marginal utility of consumption and leisure).

## Policy analysis

Firstly, we address the optimal policy in the case of autarky when shocks are purely idiosyncratic. Table 4.3 summarises the fluctuations in consumption and the tax rate in autarky for the three degrees of risk-aversion which we analyse. The metric used is the difference between the expected values of consumption and the tax rate when the government expenditure is high and when it is low. We also present a normalisation relative to the steady state values.<sup>17</sup> The governments increase their debt in periods of adverse shock realisations, but the changes in government debt are quantitatively small (around 0.1% of GDP), and the tax rate and consumption change considerably depending on the government expenditure realisation. In the case of log utility of consumption, the difference in the expected consumption between the high and the low shock realisations is 2.5% of steady state consumption, and the difference in the tax rate is 14% of steady-state tax rate. With increasing risk aversion, the differences in consumption and the tax rate between shock realisations are reduced.

Table 4.3 Fluctuations in consumption and tax rate (autarky or purely aggregate shocks)

$\sigma$	1 (log)	2	3
$\mathbb{E}[c g = g_H] - \mathbb{E}[c g = g_L]$	-0.0466	-0.0323	-0.0248
$\mathbb{E}[\tau g = g_H] - \mathbb{E}[\tau g = g_L]$	0.035	0.032	0.031
$(\mathbb{E}[c g = g_H] - \mathbb{E}[c g = g_L])/c_{ss}$	-2.5%	-1.7%	-1.3%
$(\mathbb{E}[\tau g = g_H] - \mathbb{E}[\tau g = g_L])/\tau_{ss}$	14%	12.92%	12.4%

We next address the partial union results. Figures 4.1 and 4.2 show the differences in policy and allocation between the autarky setting and partial union for preferences log in consumption, for a particular sequence of expenditure shocks. Table 4.4 shows that the households are able to achieve a much smoother consumption profile than before. For the case of log preferences ( $\sigma = 1$ ), the difference in expected consumption between the two shock realisations decreases from 2.5% percent of steady-state consumption to 0.11%. On average, the difference in the tax rate between the shock realisations is as high as in autarky. Table 4.5 shows the average changes in debt relative to GDP,<sup>18</sup> and Appendix C2 contains the

<sup>17</sup>These expectations were approximated by Monte Carlo simulation. We used 10,000 simulations of length 150 periods.

<sup>18</sup>These results were obtained using 1,000 simulations of length 50, and averaged over both countries.

full portfolios for two different normalisations. The partial union setting makes it possible to achieve close to full consumption smoothing. However, the fluctuations in labour are increased relative to autarky. In any period, it is optimal for the households which are in the low-expenditure country (and pay lower taxes) to work more, and lend resources to the households in the country with the higher expenditures and taxes. This is consistent with our intuition about the role of the no-arbitrage constraint, which implies that cross-country differences in consumption are very persistent and therefore costly from the social planner's perspective. The instantaneous utility of a country is actually lower when its expenditure is low, because its households work more and the consumption in both countries is almost identical. The results are similar for greater degrees of risk-aversion.

Because markets are incomplete, both the consumption and the tax rates contain a random walk component as in AMSS, i.e. are persistently affected by shock realisations. However, fluctuations in labour supply absorb the majority of the shocks, and consumption and tax policy in both countries remain close to each other for a relatively long time period.

When a country experiences an adverse expenditure shock ( $g = g_H$ ), its government increases its debt slightly (on average by 0.179% of GDP). In the country with the low shock realisation, the government decreases its debt (0.357% of GDP). The changes in government debt are small relative to the government expenditure shocks, which is reflected in the result that the average change in the tax rate does not decrease relative to autarky. The households in the two countries are able to achieve a smooth consumption profile by borrowing from each other. The borrowing/saving by households results in a significant change in the net external debt of a country every period (2.73% of GDP on average in either direction) compared to the small change in government debt. The shocks have a significant and persistent effect on the net external debt. Figure 4.3 shows the sample quantiles of net external debt relative to GDP over 1000 simulations, for  $\sigma = 1$ . This result highlights the importance of using a global solution method.

An important result is that the optimal policy involves quantitatively small changes in government debt compared to the significant changes in the net external debt of a country and that taxes fluctuate approximately as much as in autarky. The vast majority of borrowing

Fig. 4.1 Autarky vs. partial union (policy)

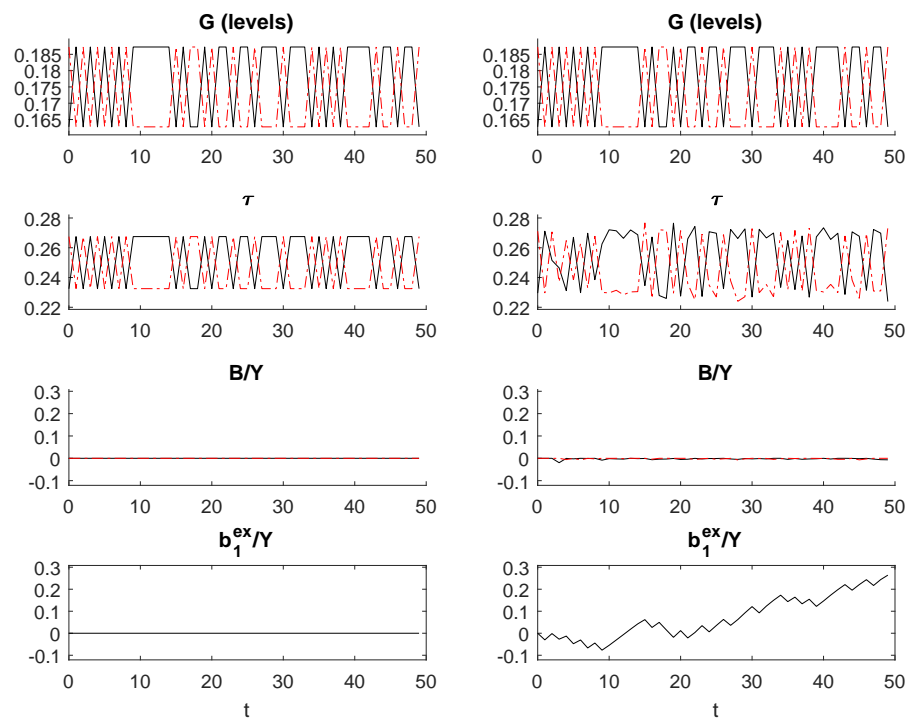




Fig. 4.2 Autarky vs. partial union (allocation)

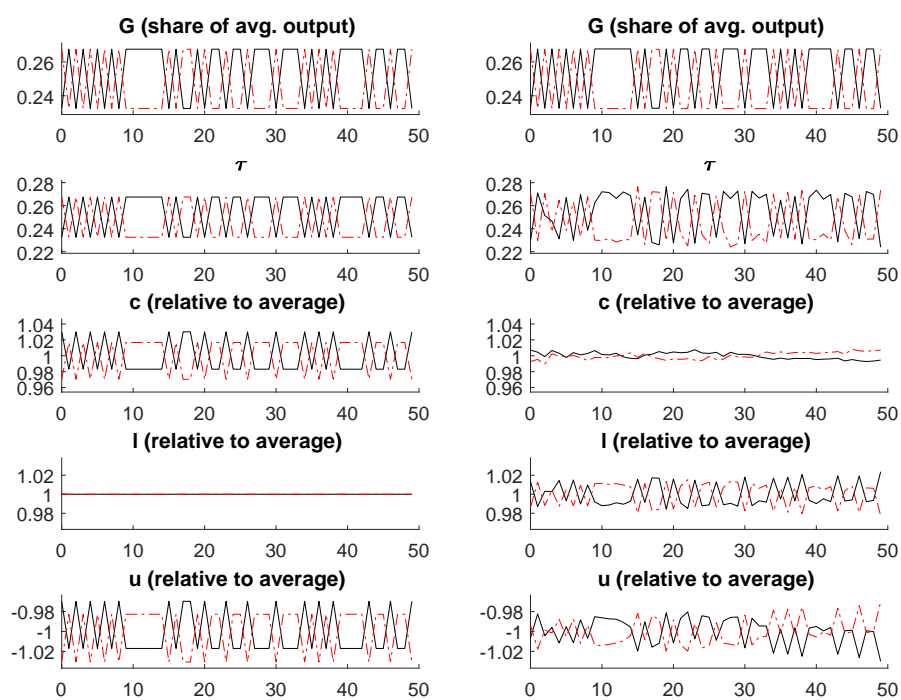
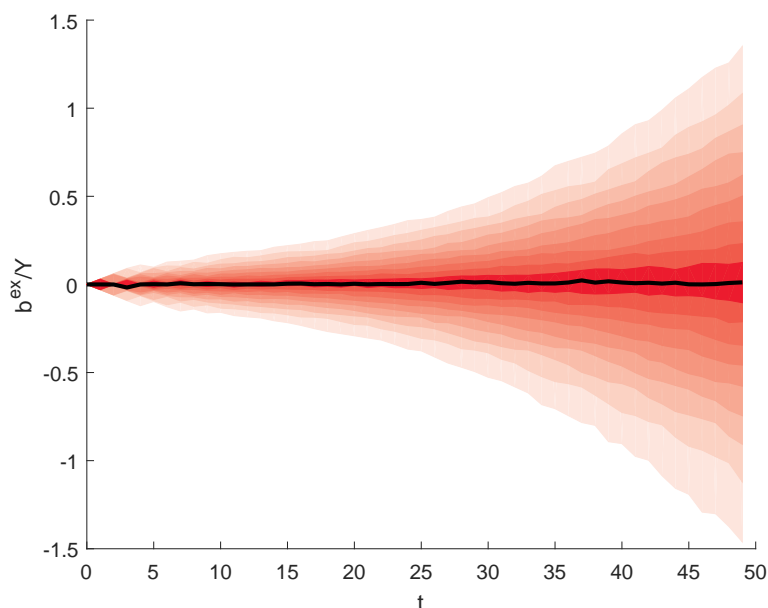


Fig. 4.3 Net external debt/GDP of country 1 quantiles in 0.05 steps (partial union)



is done between households.<sup>19</sup> One might expect that a better policy would be for the governments to borrow directly from each other, in order to reduce the distortionary fluctuations in the tax rate. However, this policy is in fact not optimal. If the governments significantly borrow and save in response to shocks (which are purely idiosyncratic), this implies that their spending needs in the future will be persistently different. This would lead to a greater expected difference in the tax rates of the two countries in the future, which is suboptimal due to convex distortion. This result relies on the assumption of bounded asset profiles, which prevents the governments from running Ponzi schemes (Bhandari et al. (2013)). These would allow the governments to borrow from each other without ever increasing the distortionary tax rate in the future.<sup>20</sup>

Finally, we address the full union results. In the full union, the set of allocations which the social planner can achieve is affected only by the aggregate expenditure of the

<sup>19</sup>The households borrow from each other through the use of government-issued bonds, and the debt of governments stays close to zero (see Figures C.1 and C.2 in Appendix C2).

<sup>20</sup>A Ponzi scheme on part of the governments is prevented by our solution method (value function iteration on a discrete grid). Our choice of grid boundaries imposes an implicit limit on the borrowing of each government of around 600% of steady-state output for the results presented in this section. These limits are never approached as the governments' debt remains close to zero, and changing the grid boundaries does not affect the result of the very limited government borrowing (as long as grids are sufficiently dense to be able to produce an accurate solution).

Table 4.4 Fluctuations in consumption and tax rate in the partial union

$\sigma$	1 (log)	2	3
$(\mathbb{E}[c g = g_H] - \mathbb{E}[c g = g_L])/c_{ss}$	-0.11%	-0.058%	-0.067 %
$(\mathbb{E}[\tau g = g_H] - \mathbb{E}[\tau g = g_L])/\tau_{ss}$	14%	13.2 %	12.5%

Table 4.5 Fluctuations in assets in the partial union

$\sigma$	1 (log)	2	3
$\mathbb{E}[\Delta b/Y g = g_L]$	-0.36 %	-0.41 %	-0.33 %
$\mathbb{E}[\Delta b/Y g = g_H]$	0.18 %	0.31 %	0.22 %
$\mathbb{E}[\Delta b^{ex}/Y g = g_L]$	-2.78 %	-2.76 %	- 2.54 %
$\mathbb{E}[\Delta b^{ex}/Y g = g_H]$	2.68 %	2.68%	2.92 %

two governments. When shocks are purely idiosyncratic, the aggregate expenditure has zero variance and perfect consumption smoothing can be achieved by choosing transfers optimally.<sup>21</sup> Figures 4.4 and 4.5 compare the optimal policy and allocation in the partial union and the full union.

## Welfare analysis

As we noted in the discussion of the implications of the no-arbitrage conditions, the welfare ranking of the three settings is in general not clear-cut due to the additional no-arbitrage constraint in the union setting.

In our numerical examples with ex ante identical countries and purely idiosyncratic shocks, the full union setting is the best in terms of social welfare. It allows the Ramsey planner to achieve a complete markets allocation (perfectly smooth consumption across countries and states), which is strictly better than the welfare in the other two settings.

Because the countries are ex ante identical and the initial asset holdings are zero, we expect autarky to result in the lowest social welfare. In the partial union, the ability of the countries to borrow internationally offers standard risk-sharing benefits. On the other hand,

<sup>21</sup>This setting corresponds to a large country with two heterogeneous households as in Shin (2006), household specific tax rates, and no uncertainty.

Fig. 4.4 Partial union vs. full union (policy)

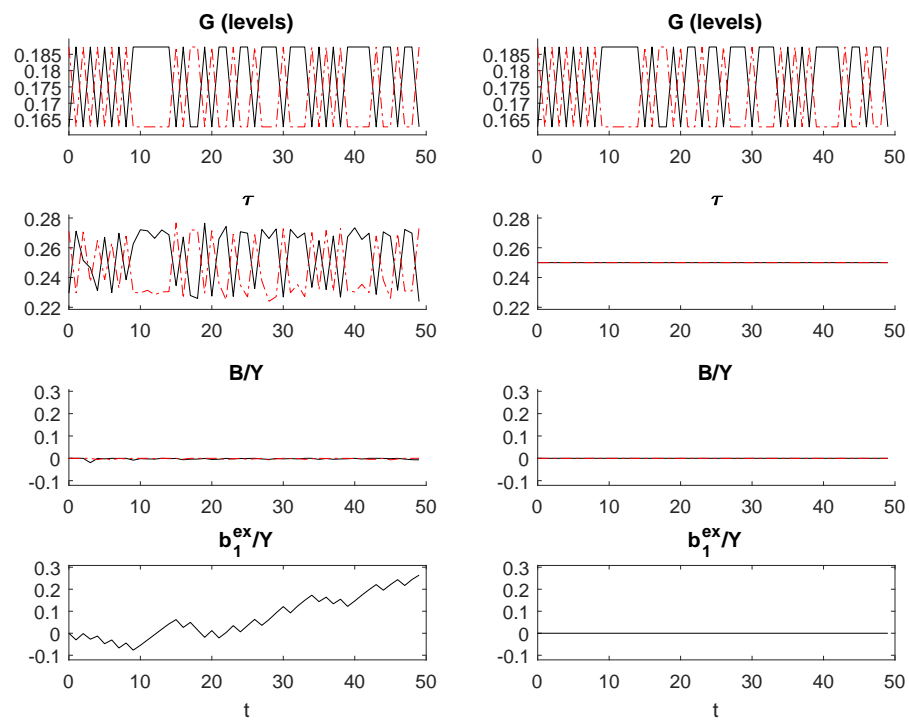
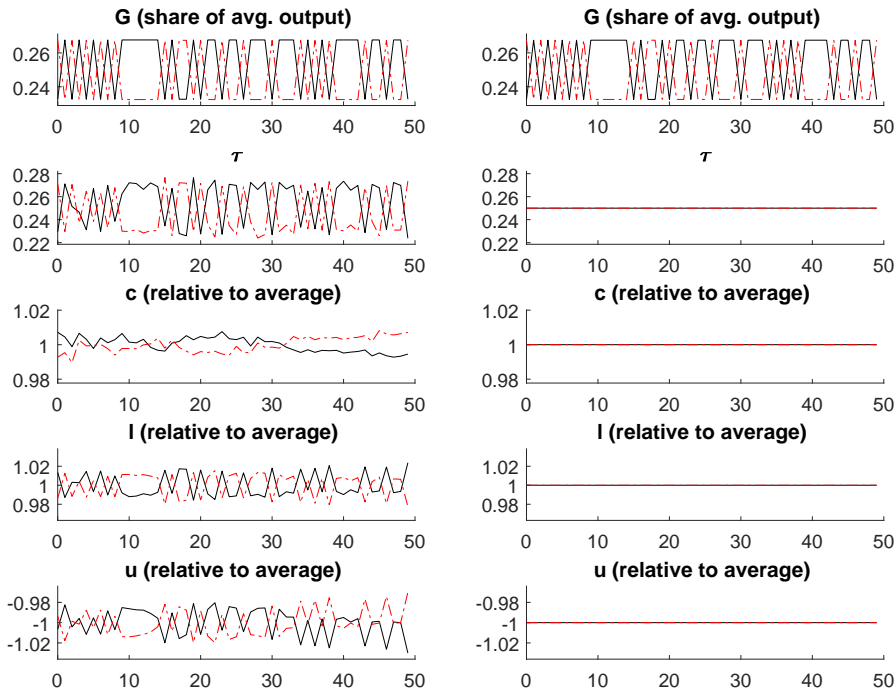


Fig. 4.5 Partial union vs. full union (allocation)



the constraint that the interest rate on the bonds of each country has to be equal does not appear to be important for welfare when there is limited heterogeneity between countries.<sup>22</sup>

We first compare welfare between the worst case (autarky) and the best case (full union), in order to find out how large the welfare gap between these settings is. We then compare the intermediate case (partial union) with the other two. Table 4.6 shows the differences in the expected social welfare between the full union and autarky settings. Table 4.7 represents the same information in terms of consumption equivalence (the percentage increase in consumption in every history in autarky).

If the consumption of households in autarky is increased by approximately 0.033%, they are as well off as in the full union. The consumption equivalence results are very similar for the studied range of risk aversion. If households are more risk averse, then the cost of fluctuating consumption in terms of social welfare level is greater. However, the percentage increase in consumption which is needed to bring about the increase in welfare stays approximately the same.

Table 4.6 Welfare gains relative to autarky (SWF difference)

$\sigma$	full union
1 (log)	0.0067
2	0.0128
3	0.0251

Table 4.7 Welfare gains relative to autarky (consumption equivalence)

$\sigma$	full union
1 (log)	0.0335 %
2	0.0335 %
3	0.0337 %

Even though the differences in terms of policy are substantial, the welfare gap between the worst and the best setting is quantitatively small. In the partial union, the fluctuations in consumption decrease relative to autarky, which suggests that the risk-averse households

<sup>22</sup>In a similar setting with ex ante identical countries and purely idiosyncratic shocks but bonds of two maturities, Equiza-Goni et al. (2016) show that the no-arbitrage constraint results in welfare lower than in autarky, because it prevents the social planner from achieving the complete markets outcome by fiscal hedging.

should be better off. In fact, consumption is almost as smooth as in the full union. However, labour supply now fluctuates more and the fluctuations in the instantaneous utility are not reduced significantly (see Figure 4.2). In the case of log utility of consumption, the percentage increase in consumption in autarky needed to achieve the same welfare as in the partial union is between 0.005% and 0.012%, using different simulation parameters.<sup>23</sup>

To summarise our numerical findings, the welfare of all three settings is quantitatively very close. International borrowing with incomplete markets (one-period bonds only) results in welfare which is only very slightly greater than in autarky, and relatively closer to autarky than to the full union (complete markets) benchmark.

## 4.5 Conclusion

In this article, we study the implications of extending the closed economy model of Aiyagari et al. (2002) into an open economy setting with frictionless capital markets. We solve the multi-country model of optimal fiscal policy using a global solution method. We quantitatively analyse the optimal policy and welfare in three different settings, and shocks either purely idiosyncratic, or purely aggregate.

The results in the partial union are similar in many respects to the closed economy results of AMSS, but there are important differences. In particular, consumption is much smoother than in the closed economy, but labour fluctuates more. Our results also highlight the importance of avoiding a large ex post asymmetry in the debt of the governments, which would lead to a greater tax distortion in the long run. The vast majority of international borrowing is done between households.

As a direction for our near future research, we plan to solve the model for different calibrations, in order to study the impact of ex ante heterogeneity between countries on the

---

<sup>23</sup>The parameters are the number of series, and parameters related to the accuracy of simulation. Due to the presence of the constraint which equates the (expected) growth rates of marginal utilities in the partial union, a very accurate solution is needed every period because of error propagation. If an optimisation subroutine makes even a minor mistake in a period, this error is propagated into future periods. In autarky, this constraint is not present, and the error propagation in simulation is much more benign. Narrowing this range down further is a topic for future work.

optimal policy in a partial union, particularly in the initial asset holdings, in order to identify the cases in which the no-arbitrage constraint is especially important for policy and welfare.

As the empirical evidence in Equiza-Goni et al. (2017) suggests, market incompleteness plays an important role in debt management across different EU countries. We believe that this third paper has helped us to develop a valid and powerful numerical tool that will allow us to extend our analysis to richer models. For this reason we consider this as a first step towards the solution of models that will incorporate monetary and further fiscal tools such as inflation and multiple maturities. These extensions will allow us to rank optimal policies and to describe how different policy mixes affect welfare. The results will then help us to contribute to the current European debate about institutions and fiscal policies in the context of a more or less integrated European fiscal union.



# References

- Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of labor economics*, 4:1043–1171.
- Acemoglu, D. and Restrepo, P. (2016). The race between machine and man: Implications of technology for growth, factor shares and employment. Working Paper 22252, National Bureau of Economic Research.
- Acemoglu, D. and Restrepo, P. (2017). Robots and jobs: Evidence from us labor markets. Working Paper 23285, National Bureau of Economic Research.
- Aiyagari, S. R., Marcet, A., Sargent, T. J., and Seppala, J. (2002). Optimal taxation without state-contingent debt. *Journal of Political Economy*, 110(6):pp. 1220–1254.
- Aiyagari, S. R. and McGrattan, E. R. (1998). The optimum quantity of debt. *Journal of Monetary Economics*, 42(3):447–469.
- Alesina, A. and Perotti, R. (1994). The political economy of budget deficits. Working Paper 4637, National Bureau of Economic Research.
- Angeletos, G.-M. (2002). Fiscal policy with noncontingent debt and the optimal maturity structure. *The Quarterly Journal of Economics*, 117(3):1105–1131.
- Aruoba, S. B., Fernández-Villaverde, J., and Rubio-Ramírez, J. F. (2006). Comparing solution methods for dynamic equilibrium economies. *Journal of Economic dynamics and Control*, 30(12):2477–2508.
- Atkinson, A. B. (2008). *The changing distribution of earnings in OECD countries*. Oxford University Press on Demand.
- Autor, D. (2015a). Paradox of abundance: Automation anxiety returns. *Performance and Progress: Essays on Capitalism, Business, and Society*.
- Autor, D. (2015b). Why are there still so many jobs? the history and future of workplace automation. *The Journal of Economic Perspectives*, 29(3):3–30.
- Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the other 99 percent. *Science*, 344(6186):843–851.
- Barro, R. J. (1979). On the determination of the public debt. *The Journal of Political Economy*, 87:940–971.

- Bassetto, M. (2014). Optimal fiscal policy with heterogeneous agents. *Quantitative Economics*, 5(3):675–704.
- Benigno, P. (2004). Optimal monetary policy in a currency area. *Journal of international economics*, 63(2):293–320.
- Benigno, P. and Woodford, M. (2012). Linear-quadratic approximation of optimal policy problems. *Journal of Economic Theory*, 147(1):1–42.
- Bertsekas, D. (1982). Distributed dynamic programming. *IEEE transactions on Automatic Control*, 27(3):610–616.
- Bertsekas, D. (2012). *Dynamic programming and optimal control*, volume 2: Approximate dynamic programming. Athena Scientific Belmont, MA.
- Bhandari, A., Evans, D., Golosov, M., and Sargent, T. (2017). Public debt in economies with heterogeneous agents. *The Journal of Monetary Economics* (forthcoming).
- Bhandari, A., Evans, D., Golosov, M., and Sargent, T. J. (2013). Taxes, debts, and redistributions with aggregate shocks. Technical Report No. 19470, National Bureau of Economic Research.
- Bhandari, A., Evans, D., Golosov, M., and Sargent, T. J. (2016). Fiscal policy and debt management with incomplete markets. *Working paper*.
- Bloom, D. E., Canning, D., and Fink, G. (2010). Implications of population ageing for economic growth. *Oxford review of economic policy*, 26(4):583–612.
- Boyan, J. and Moore, A. W. (1995). Generalization in reinforcement learning: Safely approximating the value function. *Advances in neural information processing systems*, pages 369–376.
- Brynjolfsson, E. and McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.
- Buera, F. and Nicolini, J. P. (2004). Optimal maturity of government debt without state contingent bonds. *Journal of Monetary Economics*, 51(3):531–554.
- Cai, Y. and Judd, K. L. (2014). Advances in numerical dynamic programming and new applications. *Handbook of computational economics*, 3:479–516.
- Cai, Y., Judd, K. L., Lontzek, T. S., Michelangeli, V., and Su, C.-L. (2013). Nonlinear programming method for dynamic programming. Technical report, National Bureau of Economic Research.
- Cai, Y., Judd, K. L., Thain, G., and Wright, S. J. (2015). Solving dynamic programming problems on a computational grid. *Computational Economics*, 45(2):261–284.
- Cole, H. L. and Obstfeld, M. (1991). Commodity trade and international risk sharing: How much do financial markets matter? *Journal of monetary economics*, 28(1):3–24.
- Davis, C. (2009). *Who Pays? A Distributional Analysis of the Tax Systems in All 50 States*. Institute on Taxation and Economic Policy.

- Debortoli, D., Nunes, R., and Yared, P. (2017). Optimal time-consistent government debt maturity. *The Quarterly Journal of Economics*, 132(1):55–102.
- Den Haan, W. J. and Marcet, A. (1990). Solving the stochastic growth model by parameterizing expectations. *Journal of Business & Economic Statistics*, 8(1):31–34.
- Equiza-Goni, J., Faraglia, E., and Oikonomou, R. (2016). Union debt management. *CEPR discussion paper*, (DP11181).
- Equiza-Goni, J., Faraglia, E., and Oikonomou, R. (2017). Union debt management. *mimeo*.
- Faraglia, E., Marcet, A., Oikonomou, R., and Scott, A. (2014a). Government debt management: The long and the short of it. *CEPR discussion paper*, (DP10281).
- Faraglia, E., Marcet, A., Oikonomou, R., and Scott, A. (2014b). Optimal fiscal policy problems under complete and incomplete financial markets: A numerical toolkit.
- Faraglia, E., Marcet, A., and Scott, A. (2008). Fiscal insurance and debt management in oecd economies. *The Economic Journal*, 118(527):363–386.
- Faraglia, E., Marcet, A., and Scott, A. (2010). In search of a theory of debt management. *Journal of Monetary Economics*, 57(7):821–836.
- Farhi, E. (2010). Capital taxation and ownership when markets are incomplete. *Journal of Political Economy*, 118(5):pp. 908–948.
- Farhi, E. and Werning, I. (2007). Inequality and social discounting. *Journal of political economy*, 115(3):365–402.
- Farhi, E. and Werning, I. (2017). Fiscal unions.
- Ferrero, A. (2009). Fiscal and monetary rules for a currency union. *Journal of International Economics*, 77(1):1–10.
- Gali, J. and Monacelli, T. (2008). Optimal monetary and fiscal policy in a currency union. *Journal of international economics*, 76(1):116–132.
- Golosov, M. and Sargent, T. J. (2012). Taxation, redistribution, and debt with aggregate shocks.
- Grossman, G. and Helpman, E. (2001). Intergenerational redistribution with short-lived governments. *The Economic Journal*, 108(450):1299–1329.
- Haykin, S. (2009). *Neural networks and learning machines*, volume 3. Pearson Education Upper Saddle River.
- Heer, B. and Maussner, A. (2009). *Dynamic general equilibrium modeling: computational methods and applications*. Springer Science & Business Media.
- Heer, B. and Maussner, A. (2011). Value function iteration as a solution method for the ramsey model. *Jahrbücher für Nationalökonomie und Statistik*, 231(4):494–515.

- Hornstein, A., Krusell, P., and Violante, G. L. (2005). The effects of technical change on labor market inequalities. *Handbook of economic growth*, 1:1275–1370.
- Judd, K. (1998). *Numerical Methods in Economics*. MIT Press.
- Kim, J., Kim, S. H., and Levin, A. (2003). Patience, persistence, and welfare costs of incomplete markets in open economies. *Journal of International Economics*, 61(2):385–396.
- Kollmann, R., Maliar, S., Malin, B. A., and Pichler, P. (2011). Comparison of solutions to the multi-country real business cycle model. *Journal of Economic Dynamics and Control*, 35(2):186–202.
- Kydland, F. E. and Prescott, E. C. (1980). Dynamic optimal taxation, rational expectations and optimal control. *Journal of Economic Dynamics and Control*, 2:79–91.
- Lemieux, T. (2008). The changing nature of wage inequality. *Journal of Population Economics*, 21(1):21–48.
- Lewis, K. K. (1996). What can explain the apparent lack of international consumption risk sharing? *Journal of Political Economy*, 104(2):267–297.
- Lewis, K. K. (2000). Why do stocks and consumption imply such different gains from international risk sharing? *Journal of international Economics*, 52(1):1–35.
- Lucas, R. E. and Stokey, N. L. (1983). Optimal fiscal and monetary policy in an economy without capital. *Journal of Monetary Economics*, 12(1):55 – 93.
- Metcalf, M., Reid, J., and Cohen, M. (2011). *Modern Fortran Explained*. Oxford University Press.
- Miranda, M. J. and Fackler, P. L. (2004). *Applied computational economics and finance*. MIT press.
- Mueller, D. C. (2003). *Public Choice III*. Cambridge University Press.
- Munos, R. and Szepesvári, C. (2008). Finite-time bounds for fitted value iteration. *Journal of Machine Learning Research*, 9(May):815–857.
- Nguyen, A. (2012). Lecture notes in machine learning (cs229), stanford university, available at <http://cs229.stanford.edu/materials.html>.
- Niepelt, D. (2004). Tax smoothing versus tax shifting. *Review of Economic Dynamics*, 7(1):27 – 51.
- Peterman, W. B. (2014). Reconciling micro and macro estimates of the frisch labor supply elasticity: A sensitivity analysis. Technical report, Working Paper.
- Piguillem, F. and Schneider, A. L. (2013). Heterogeneous labor skills, the median voter and labor taxes. *Review of Economic Dynamics*, 16(2).
- Poli, R., Kennedy, J., and Blackwell, T. (2007). Particle swarm optimization. *Swarm intelligence*, 1(1):33–57.

- Röhrs, S. and Winter, C. (2017). Reducing government debt in the presence of inequality. *Journal of Economic Dynamics and Control*.
- Schreiber, A. J. E. (2016). *Dynamic programming with radial basis functions and Shepard's method*. Phd dissertation, Technische Universität München.
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524. ACM.
- Shin, Y. (2006). Ramsey meets bewley: Optimal government financing with incomplete markets. *Unpublished manuscript, Washington University in St. Louis*.
- Stachurski, J. (2008). Continuous state dynamic programming via nonexpansive approximation. *Computational Economics*, 31(2):141–160.
- Tabellini, G. (1991). The politics of intergenerational redistribution. *Journal of Political Economy*, 99(2):335–57.
- Van Wincoop, E. (1994). Welfare gains from international risksharing. *Journal of Monetary Economics*, 34(2):175–200.
- Van Wincoop, E. (1999). How big are potential welfare gains from international risksharing? *Journal of International Economics*, 47(1):109–135.
- Werning, I. (2007). Optimal fiscal policy with redistribution. *The Quarterly Journal of Economics*, 122(3):925–967.



# Appendix A

## Chapter 2 appendices

### Appendix A1: Ramsey problem and dimensionality reduction

This appendix contains derivation of the constraints in the sequential Ramsey problem of section 2.3.3, and in its recursive formulation. As a first step, the first-order conditions of agent  $i$ 's utility maximisation problem are derived.

Recall that  $s^t = (s_0, \dots, s_t)$  denotes a history of government expenditure shocks up and until time  $t$ .  $\Pr(s^t|s^k)$  is the probability of history  $s^t$  conditional on the realisation of history  $s^k$  for  $t > k$ .  $\sum_{s^t}$  is summation over all possible histories at time  $t$ , and  $\sum_{s^t|s^k}$  summation over all histories  $s^t$  which can occur after  $s^k$ , where  $t > k$ .

Every agent at time  $t = 0$  chooses a history-contingent plan for all periods  $t$  and histories  $s^t$ , taking prices and government policy as given. Because of assumptions made about the environment, we can assume that agents always choose interior consumption and labour supply, and that the agents' budget constraints are satisfied with equality<sup>1</sup>. I assume that there are no binding borrowing constraints. The problem of agent  $i$  is then

$$\max_{\{c_{i,t}(s^t), l_{i,t}(s^t), b_{i,t}(s^t)\}_{t=0, \forall s^t}^{\infty}} \mathbb{E}_0 \left[ \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right]$$

---

<sup>1</sup>The interior solution for labour supply does not follow from properties of the utility function used in the numerical example, because marginal utility of leisure is not infinity at the maximum feasible labour. With reasonable parameter values, and interior solution is always chosen.

subject to

$$c_{i,t}(s^t) + b_{i,t}(s^t) = (1 - \tau_t(s^t))\theta_{i,t}l_{i,t}(s^t) + R_{t-1}(s^{t-1})b_{i,t-1}(s^{t-1}), \forall t, s^t$$

The borrowing constraints depend on time but not on history, and are in levels. The Lagrangian is

$$\begin{aligned} L = & \sum_{t=0}^{\infty} \sum_{s^t} \beta^t U^i(c_{i,t}(s^t), l_{i,t}(s^t)) \Pr(s^t | s_0) \\ & + \sum_{t=0}^{\infty} \sum_{s^t} \beta^t \lambda_t(s^t) [(1 - \tau_t(s^t))\theta_{i,t}l_{i,t}(s^t) + R_{t-1}(s^{t-1})b_{i,t-1}(s^{t-1}) - c_{i,t}(s^t) - b_{i,t}(s^t)] \end{aligned}$$

where  $\beta^t \lambda_t(s^t)$  are Lagrange multipliers on the budget constraint. The first-order conditions  $\forall t, s^t$  are

$$U_c^i(c_{i,t}(s^t), l_{i,t}(s^t)) \Pr(s^t | s_0) - \lambda_t(s^t) = 0 \quad (\text{A.1})$$

$$U_l^i(c_{i,t}(s^t), l_{i,t}(s^t)) \Pr(s^t | s_0) + \lambda_t(s^t)(1 - \tau_t(s^t))\theta_{i,t} = 0 \quad (\text{A.2})$$

$$-\lambda_t(s^t) + \beta R_t(s^t) \sum_{s^{t+1}|s^t} \lambda_{t+1}(s^{t+1}) = 0 \quad (\text{A.3})$$

$$(\text{A.4})$$

Combining these first-order-conditions, we obtain the intratemporal and intertemporal Euler equations, or conditions for tax policy and gross rate of returns

$$1 - \tau_t(s^t) = - \frac{U_l^i(c_{i,t}(s^t), l_{i,t}(s^t))}{\theta_{i,t} U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))} \quad (\text{A.5})$$

$$R_t(s^t) = \frac{U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))}{\beta \mathbb{E}_t[U_c^i(c_{i,t+1}, l_{i,t+1})]} \quad (\text{A.6})$$

We can use these equations to eliminate prices and tax policy in the agent's budget constraint to solve for the optimal allocation directly. We can combine equations (A.5), (A.6) and the agent's budget constraint (2.10) to derive implementability constraints, which hold  $\forall i, t, s^t$  at an interior solution:

$$c_{i,t}(s^t) + b_{i,t}(s^t) = - \frac{U_l^i(c_{i,t}(s^t), l_{i,t}(s^t))}{U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))} l_{i,t}(s^t) + \frac{U_c^i(c_{i,t-1}(s^{t-1}), l_{i,t-1}(s^{t-1}))}{\beta \mathbb{E}_{t-1}[U_c^i(c_{i,t}, l_{i,t})]} b_{i,t-1}(s^{t-1}) \quad (\text{A.7})$$



The other constraints that follow from equating taxes and prices across agents using equations (A.5) and (A.6) are,  $\forall i, j, t, s^t$ :

$$\frac{U_l^i(c_{i,t}(s^t), l_{i,t}(s^t))}{\theta_{i,t} U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))} = \frac{U_l^j(c_{j,t}(s^t), l_{j,t}(s^t))}{\theta_{j,t} U_c^j(c_{j,t}(s^t), l_{j,t}(s^t))} \quad (\text{A.8})$$

$$\frac{U_c^i(c_{i,t}(s^t), l_{i,t}(s^t))}{\mathbb{E}_t[U_c^i(c_{i,t+1}, l_{i,t+1})]} = \frac{U_c^j(c_{j,t}(s^t), l_{j,t}(s^t))}{\mathbb{E}_t[U_c^j(c_{j,t+1}, l_{j,t+1})]} \quad (\text{A.9})$$

We can recover the government's assets by using the market-clearing condition in the asset market (2.12).

When we solve the Ramsey problem, we can ignore the government's budget constraint because by Walras' law, it is redundant once we impose the budget constraints of all agents (2.10), the aggregate resource constraint (2.5), and the market-clearing condition in asset market (2.12). Therefore, to find the Ramsey allocation, we can solve the problem of maximising (2.13) with respect to  $\{c_{i,t}(s^t), l_{i,t}(s^t), b_{i,t}(s^t)\}_{t=0, i=1, \forall s^t}^{\infty, I}$ , subject to the aggregate resource constraint (2.5), and the implementability conditions (A.7), (A.8), and (A.9).

The recursive formulation is a variation on Shin (2006) and Bhandari et al. (2013) with the addition of state variable  $t$  with exogenous law of motion, and a functional dependence of labour productivities of the agents on this state. Because the functions are assumed to become constant for some  $\mathcal{T}$ , we can choose a compact state space for  $t$ .

## Reduction of dimensionality

The number of state variables in the problem of section 2.3.4 is rather large which creates difficulties in finding a solution using value function iteration on a discrete grid. We can find a recursive formulation with one fewer state variable along the lines of Shin (2006). This section gives one such formulation for the special case of two agents which is the case that I focus on in applications of the model. It can be generalised to an arbitrary number of agents (see Bhandari et al. (2013)).

With two agents the state vector is  $(b_1, b_2, u_1, u_2, t, s_-)$ , where  $b_i$  is last-period asset holdings of agent  $i$ ,  $u_i$  is  $i$ 's last-period marginal utility of consumption,  $t$  is 'time' governing productivity trends, and  $s_i$  is last-period shock realisation.

Now define  $a_i = \frac{1}{\beta} u_i b_i$  for  $i = 1, 2$ , and  $\rho = \frac{u_1}{u_2}$ . The new state vector is  $(a_1, a_2, \rho, t, s_-)$ . The Bellman equation for periods  $t \geq 1$  is

$$V(a_1, a_2, \rho, t, s_-) =$$

$$\max_{[c_i(s), l_i(s), a_i'(s)]_{i,s}} \sum_s Pr(s|s_-) \left[ \pi_1 \alpha_1 U^1(c_1(s), l_1(s)) + \pi_2 \alpha_2 U^2(c_2(s), l_2(s)) + \beta V(a_1'(s), a_2'(s), \frac{U_c^1(s)}{U_c^2(s)}, t', s) \right]$$

subject to

$$c_1(s) + \frac{a_1'(s)}{\beta U_c^1(s)} + \frac{U_l^1(s)}{U_c^1(s)} l_1(s) = \frac{a_1}{\mathbb{E}_{s_-}[U_c^1]}, \forall s \quad (\text{A.10})$$

$$c_2(s) + \frac{a_2'(s)}{\beta U_c^2(s)} + \frac{U_l^2(s)}{U_c^2(s)} l_2(s) = \frac{a_2}{\mathbb{E}_{s_-}[U_c^2]}, \forall s \quad (\text{A.11})$$

$$\rho = \frac{\mathbb{E}_{s_-}[U_c^1]}{\mathbb{E}_{s_-}[U_c^2]} \quad (\text{A.12})$$

$$\frac{U_l^1(s)}{\theta_1 U_c^1(s)} = \frac{U_l^2(s)}{\theta_2 U_c^2(s)}, \forall s \quad (\text{A.13})$$

$$\sum_{i=1}^I \pi_i c_i(s) + g = \sum_{i=1}^I \pi_i \theta_i l_i(s), \forall s \quad (\text{A.14})$$

$$\theta_i = \theta_i(t) \forall i \quad (\text{A.15})$$

$$t' = \min\{t + 1, \mathcal{T}\} \quad (\text{A.16})$$

$$g = \bar{g}s, \forall s \quad (\text{A.17})$$

$$\bar{g} = k \sum_{i=1}^I \pi_i \theta_i \quad (\text{A.18})$$

As is the case with the original formulation, the period 0 problem is solved separately as the last-period states (which depend on marginal utilities) are undefined. The problem is a slight variation on the first-period problem given in section 2.3.4. This formulation assumes that the agents' borrowing constraints do not bind.

## Appendix A2: Calibration details

I describe calibration for the utility function

$$U^i(c_{i,t}, l_{i,t}) = \log(c_{i,t}) - B \frac{l_{i,t}^{1+\gamma}}{1+\gamma}.$$

Discount factor  $\beta$  is chosen such that the interest rate in steady state<sup>2</sup> matches a given target:  $R_{tgt}\beta = 1$ .

Parameter  $\gamma$  is chosen to target Frisch elasticity of labour supply  $\eta_{tgt}$  as  $\gamma = 1/\eta_{tgt}$ . Parameter  $B$  is then chosen to target a steady-state labour  $l_{tgt}$  according to  $B = l_{tgt}^{-1-\gamma}$ .

There are a number of ways in which to proceed, depending on what data we want to use. In general, the productivities can follow any functions that eventually become constant, including non-monotonic ones. I describe the calibration used in the application to the increase in inequality since the 1980s. I use a constant growth rate specification mainly because it is easy to relate to the two-period model and interpret the results.

The calibration target is the 90-10 earnings ratio at two points in time, call these  $t = 0$  and  $t = \mathcal{T}$  (usually one period corresponds to one year). With two agents and  $\pi_i \in [0.1, 0.9]$ , this is simply the ratio  $l_1\theta_1/l_2\theta_2$ , as long as agent 1 has higher income. In a deterministic steady-state with no assets the labour supply is the same for both agents so the income ratio is just  $\theta_1/\theta_2$ . I normalise the productivity of agent 2 as  $\theta_2 = 1$ , and set the productivity of agent 1 in the two periods such that the ratio of productivities hits the target (in this case simply  $\theta_1 = ratio_{tgt}$ ). We then only need to compute the growth rate of agent 1's productivity such that the target period  $\mathcal{T}$  productivity is achieved.

Because of intertemporal substitution of labour due to the presence of heterogeneous growth (the faster-growing agent 1 works relatively more in later periods), the actual 90-10 earnings ratio will be lower in period  $t = 0$  (and higher in period  $t = \mathcal{T}$ ) than the target. If the intertemporal substitution is strong, we may want to adjust  $\theta_{1,0}$  upwards and the growth rate  $\xi_{\theta,1}$  downwards until the model gets closer to the target earnings ratio.

As for the target ratio and  $\mathcal{T}$ , Autor (2014) provides data on 90-10 earnings ratio in several OECD countries in 1980 and 2011. Alternative calibration strategies might involve targeting productivity of groups with different education, etc.

The adjusted government expenditure is chosen such that the steady-state tax rate equals target  $\tau_{tgt}$  according to the following expression which holds in a steady state with zero assets

$$g = \frac{(\pi_1 + \pi_2 \frac{\theta_2}{\theta_1})(\tau_{tgt} - 1)}{\frac{B}{\theta_1}(\frac{1}{B})^{\gamma/(1+\gamma)}} + \sum_i \pi_i \theta_i \left(\frac{1}{B}\right)^{1/(1+\gamma)}.$$

The actual government expenditure is related to the maximum feasible output in a period according to equation (2.6).

The numerical examples in this paper use zero initial asset holdings as the focus of the paper is on the change in policy due to trends, not the impact of varying initial asset

<sup>2</sup>When I refer to steady state in this section I mean a deterministic steady-state in which productivities remain at their initial levels, and there are no asset holdings.

holdings. It is slightly more convenient to work with zero initial assets because we can then use narrower grids and obtain a more accurate solution, but the difference is not substantial as the initial public debt in the US in 1980 was low (around 40% of GDP), and this level is surpassed in the simulations after several years with some choices of Pareto weights.

Table A.1 Summary of calibration

<b>Calibration targets</b>	
Annual interest rate	5.26%
Frisch elasticity of labour	0.5
labour supply	0.7
initial 90/10 earnings ratio	3.6
periods of growth	30
final 90/10 earnings ratio	5.1
tax rate	0.25
output decline in recession	3 %
expect. duration of recession	2.33 years
expect. duration of booms	7 years
<b>Parameters</b>	
$I$ (number of agents)	2
$(\pi_1, \pi_2)$ (mass of agents)	(0.5, 0.5)
$\beta$ (discount factor)	0.95003
$B$ (utility function parameter)	2.9155
$\gamma$ (utility function parameter)	2
$(\bar{l}_1, \bar{l}_2)$ (maximum labour)	(3, 3)
$(\theta_{1,0}, \theta_{2,0})$ (initial productivities)	(3.6, 1)
$(\xi_{\theta,1}, \xi_{\theta,2})$ (growth rates of productivities)	(0.0117, 0.0)
$(s_1, s_2)$ (expenditure shock realisations)	(0.9, 1.1)
$P_{1,1}$	0.6
$P_{2,2}$	0.85
$k$ (gov. expenditure constant)	0.058333
$(b_{1,-1}, b_{2,-1}, B_{-1})$ (initial assets)	(0, 0, 0)

# Appendix B

## Chapter 3 appendices

### Appendix B1: Definitions and Ramsey problem

This appendix contains formal statements of some definitions and the Ramsey problem which were omitted from the main text.

**Definition** (*allocation, asset profile, price system, tax policy*)

An *allocation* is a sequence  $\{c_{i,t}, l_{i,t}\}_{i,t}$ . An *asset profile* is a sequence  $\{\{b_{i,t}\}_i, B_t\}_t$ . A *price system* is a sequence  $\{R_t\}_t$  of interest rates on risk-free 1-period bonds. A *tax policy* is a sequence of tax rates on labour income and transfers  $\{\tau_t, T_t\}_t$ .

**Definition** (*competitive equilibrium*)

For given initial assets held by the agents  $\{b_{i,-1}\}_i$  and by the government ( $B_{-1}$ ), and a given tax policy  $\{\tau_t, T_t\}_t$ , a *competitive equilibrium* is an allocation  $\{c_{i,t}, l_{i,t}\}_{i,t}$ , an asset profile  $\{\{b_{i,t}\}_i, B_t\}_t$  and a price system  $\{R_t\}_t$ , such that:

1. The allocation  $\{c_{i,t}, l_{i,t}\}_{i,t}$  and assets held by agents  $\{b_{i,t}\}_{i,t}$ , maximise all agents' utility (3.5) subject to agent-specific budget constraints (3.10).
2. The resource constraint (3.6) is satisfied.
3. The government's budget constraint (3.11) is satisfied (this is redundant by Walras' law).
4. Market-clearing in the asset market (3.12) is satisfied.
5. The asset profile is bounded from below.

**Definition** (*socially optimal competitive equilibrium*)

Given the initial distribution of assets  $(\{b_{i,-1}\}_i, B_{-1})$ , a *socially optimal competitive equilibrium* is a tax policy  $\{\tau_t^*, T_t^*\}_t$ , an allocation  $\{c_{i,t}^*, l_{i,t}^*\}_{i,t}$ , an asset profile  $\{\{b_{i,t}^*\}_i, B_t^*\}_t$ , and a price system  $\{R_t^*\}_t$  such that

1. Given the initial asset holdings  $(\{b_{i,-1}\}_i, B_{-1})$  and the tax policy  $\{\tau_t^*\}_t$ , the other elements of the socially optimal competitive equilibrium  $\{c_{i,t}^*, l_{i,t}^*\}_{i,t}, \{\{b_{i,t}^*\}_i, B_t^*\}_t, \{R_t^*\}_t$  constitute a competitive equilibrium.
2. Given the initial asset holdings  $(\{b_{i,-1}\}_i, B_{-1})$ , there is no other tax policy  $\{\tau_t\}_t \neq \{\tau_t^*\}_t$ , for which there is a competitive equilibrium which yields higher value of social welfare (2.13).

The constraints of the Ramsey problem are derived similarly to the model without transfers. As noted in the main text, only the net asset positions are determined in an equilibrium<sup>1</sup>, where  $\tilde{b}_i = b_i - b_1$  is the net asset position of agent  $i$  relative to agent 1. Transfers  $T_t$  were eliminated using implementability constraint of agent 1.

The sequence version of Ramsey problem, i.e. the problem of finding a socially optimal allocation, is

$$\max_{\{c_{i,t}(s^t), l_{i,t}(s^t)\}_{t=0, \forall i, s^t}, \{\tilde{b}_{i,t}(s^t)\}_{t=0, \forall i > 1, \forall s^t}} \mathbb{E}_0 \left[ \sum_{i=1}^I \pi_i \alpha_i \sum_{t=0}^{\infty} \beta^t U^i(c_{i,t}, l_{i,t}) \right]$$

subject to

$$c_{i,t}(s^t) - c_{1,t}(s^t) + \tilde{b}_{i,t}(s^t) = \frac{U_l^1(s^t)}{U_c^1(s^t)} l_{1,t}(s^t) - \frac{U_l^i(s^t)}{U_c^i(s^t)} l_{i,t}(s^t) + \frac{U_c^i(s^{t-1})}{\beta \mathbb{E}_{t-1}[U_c^i]} \tilde{b}_{i,t-1}(s^{t-1}), \forall i > 1, t, s^t \quad (\text{B.1})$$

$$\frac{U_l^i(s^t)}{\theta_i(t, s_t) U_c^i(s^t)} = \frac{U_l^j(s^t)}{\theta_j(t, s_t) U_c^j(s^t)}, \forall i, j, t, s^t \quad (\text{B.2})$$

$$\sum_{i=1}^I \pi_i c_{i,t}(s^t) + g_t(s^t) = \sum_{i=1}^I \pi_i \theta_{i,t} l_{i,t}(s^t), \forall t, s^t \quad (\text{B.3})$$

As was the case in the model without transfers, interior solution with respect to labour supply in the agents' optimisation problems is assumed.

## Recursive Formulation of Ramsey Problem

The recursive formulation of this section is similar to the one in the model without transfers, the presentation of which in the main text contains an explanation of where the constraints come from.

<sup>1</sup>A more precise statement is given in Theorem 1 of Bhandari et al. (2013) which is unaffected by inclusion of trend productivity growth.

The state vector is  $(a, \rho, t, s_-)$ :

- $a = (a_2, \dots, a_I)$  are marginal-utility adjusted net asset positions  $a_i = \frac{1}{\beta} \tilde{b}_i U_c^i$
- $\rho = (\rho_2, \dots, \rho_I)$  are last-period ratios of marginal utilities of consumption  $\rho_i = U_c^1 / U_c^i$ .
- $t$  is 'time' in evolution of productivities  $\theta_i(t, s)$ .
- $s_-$  is the last-period shock realisation.

The choice variables are  $[c_i(s), l_i(s), a'_i(s)]$  for every possible realisation of shock  $s$  and for all agents  $i$  (with the exception of  $a'$  which is chosen for agents  $2, \dots, I$  only). The Bellman equation for periods  $t \geq 1$  is

$$V(a, \rho, t, s_-) = \max_{[c_i(s), l_i(s), a'_i(s)]_{i,s}} \sum_s Pr(s|s_-) \left[ \sum_{i=1}^I \pi_i \alpha_i U^i(c_i(s), l_i(s)) + \beta V(a'(s), \rho'(s), t', s) \right]$$

subject to

$$U_c^i(s)[c_i(s) - c_1(s)] + \beta a'_i(s) + U_l^i(s)l_i(s) - U_c^i(s) \frac{U_l^1(s)}{U_c^1(s)} l_1(s) = a_i \frac{U_c^i(s)}{\mathbb{E}_{s_-}[U_c^i]}, i = 2, \dots, I, \forall s \quad (\text{B.4})$$

$$\frac{\mathbb{E}_{s_-}[U_c^1]}{\mathbb{E}_{s_-}[U_c^i]} = \rho_i, \forall i = 2, \dots, I \quad (\text{B.5})$$

$$\frac{U_l^i(s)}{\theta_i(t, s) U_c^i(s)} = \frac{U_l^j(s)}{\theta_j(t, s) U_c^j(s)}, \forall i, j, s \quad (\text{B.6})$$

$$\sum_{i=1}^I \pi_i c_i(s) + g = \sum_{i=1}^I \pi_i \theta_i(t, s) l_i(s), \forall s \quad (\text{B.7})$$

$$\rho'_i(s) = \frac{U_c^1(s)}{U_c^i(s)}, \forall i = 2, \dots, I, \forall s \quad (\text{B.8})$$

$$t' = \min\{t + 1, T\} \quad (\text{B.9})$$

$$g = \bar{g}s, \forall s \quad (\text{B.10})$$

$$\bar{g} = k \sum_{i=1}^I \pi_i \theta_i \bar{l}_i \quad (\text{B.11})$$

As before, the period  $t = 0$  problem is solved separately

$$(c_0^*, l_0^*, \tilde{b}_0^*) = \arg \max_{\{c_{i,0}, l_{i,0}, \tilde{b}_{i,0}\}_i} \sum_{i=1}^I \pi_i \alpha_i U^i(c_0, l_0) + \beta V(a_0, \rho, 1, s_0)$$

subject to

$$\begin{aligned}
c_{i,0} - c_{1,0} + \tilde{b}_{i,0} + \frac{U_l^i(c_{i,0}, l_{i,0})}{U_c^i(c_{i,0}, l_{i,0})} l_{i,0} - \frac{U_l^1(c_{1,0}, l_{1,0})}{U_c^1(c_{1,0}, l_{1,0})} l_{1,0} &= \tilde{b}_{i,-1}, \forall i = 2, \dots, I \\
\frac{U_l^i(c_{i,0}, l_{i,0})}{\theta_i(0, s_0) U_c^i(c_{i,0}, l_{i,0})} &= \frac{U_l^j(c_{j,0}, l_{j,0})}{\theta_j(0, s_0) U_c^j(c_{j,0}, l_{j,0})}, \forall i, j \\
\sum_{i=1}^I \pi_i c_{i,0} + g_0 &= \sum_{i=1}^I \pi_i \theta_i(0, s_0) l_{i,0} \\
a_{i,0} &= \frac{1}{\beta} \tilde{b}_{i,0} U_c^i(c_{i,0}, l_{i,0}), i = 2, \dots, I \\
\rho_i(s) &= \frac{U_c^1}{U_c^i}, i = 2, \dots, I \\
g_0 &= \bar{g}_0 s_0 \\
\bar{g}_0 &= k \sum_{i=1}^I \pi_i \theta_{i,0} \bar{l}_i.
\end{aligned}$$

## Appendix B2: Further results

In this appendix I present further results relevant to the application of the model to 1980-2010 US data. Figures B.1 and B.2 show solutions for extreme choices of Pareto weights. These essentially correspond to individually optimal policies.

The case of Pareto weight on the high-skilled approaching 1 from the left (Figure B.1) results in a very inequitable outcome. The social planner uses lump sum taxes (negative transfers) to collect more revenue than is needed to finance government expenditures alone, in order to finance labour subsidies. The low-skilled work more but their productivity is so much lower that their labour earnings are lower, and the high-skilled thus disproportionately benefit from the labour subsidy. In later years the earnings inequality becomes even greater, and the social planner increases the labour subsidy (and increases lump sum taxes further). The Pareto weight on the high-skilled was capped at a level below one because if it were one, the social planner would want to set lump sum taxes so high that consumption of the low-skilled would be so close to zero that it would be very difficult to solve the model accurately<sup>2</sup>. Even with Pareto weight capped at  $\alpha_1 = 0.99$  we already obtain an extremely inequitable outcome which is not similar to policies observed in the real world, and it does not appear useful to solve the model numerically for  $\alpha_1$  even closer to one.

<sup>2</sup>If the consumption of the low-skilled is positive, the social planner can marginally increase lump sum taxes without violating the budget constraint of the low-skilled, and use the revenue to increase labour subsidy, which improves the welfare of the high-skilled.



Fig. B.1 Optimal policy over 30 years without shocks ( $\alpha_1 = 0.001, \alpha_2 = 0.999$ ). In the last row the solid lines corresponds to the high-skilled agent (1), the dashed lines to the low-skilled (2).

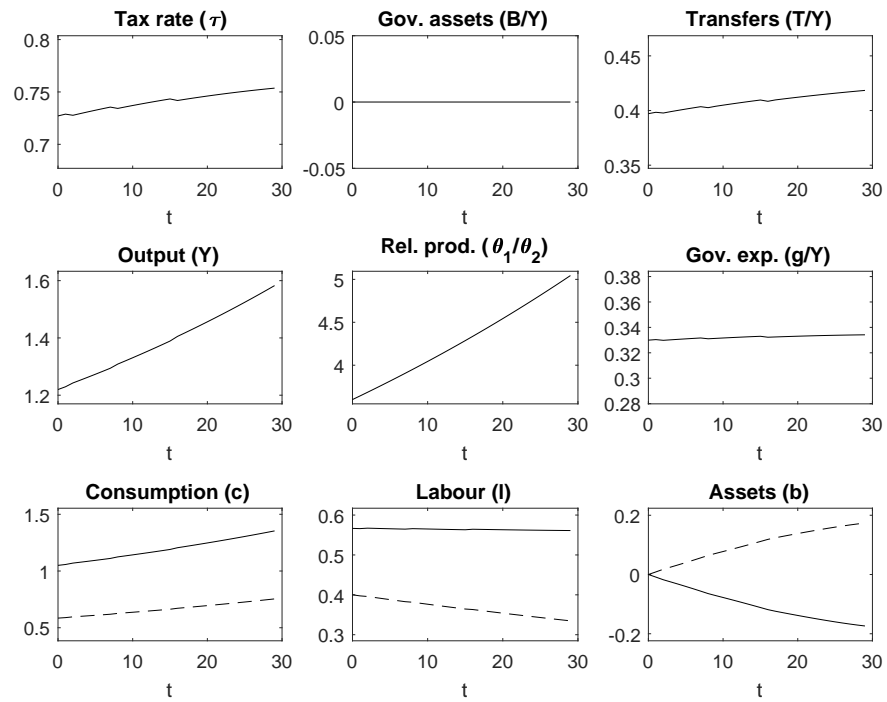
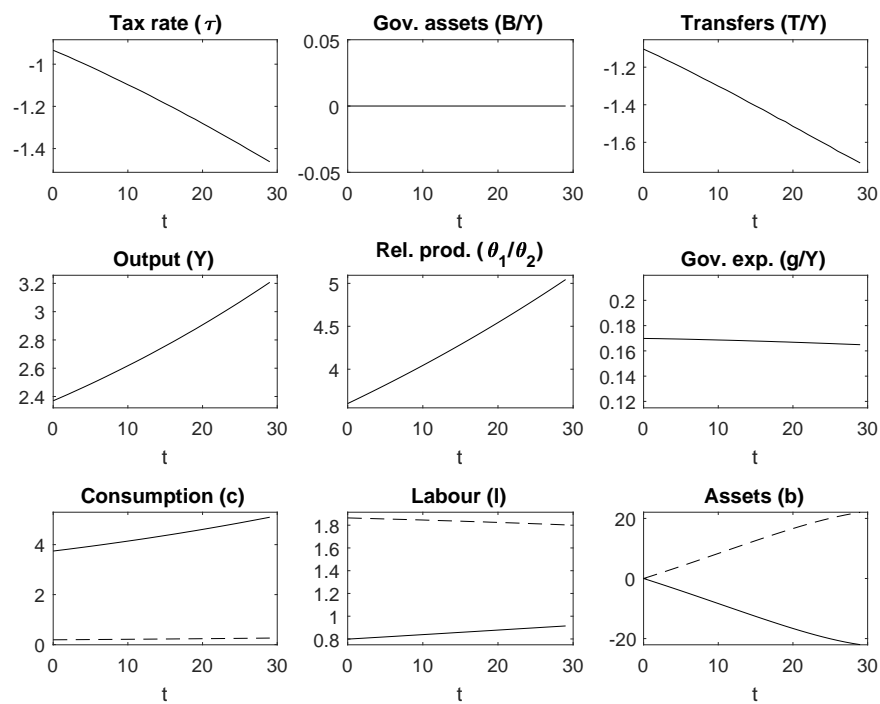


Fig. B.2 Optimal policy over 30 years without shocks ( $\alpha_1 = 0.99, \alpha_2 = 0.01$ ). In the last row the solid lines corresponds to the high-skilled agent (1), the dashed lines to the low-skilled (2).



## Appendix B3: Solution methods

This appendix contains a more extensive discussion of the solution methods. I present the fitted value iteration algorithm more formally than in the main text. I then outline the main steps of the two main algorithms: discrete value iteration, and fitted value iteration, and go into some algorithm-specific implementation details. Then I address topics common to both algorithms, such as the maximisation step in VFI at a point in state space, and various acceleration techniques (Howard's algorithm, CTS algorithm).

### Fitted value function iteration

The value function for each shock realisation is approximated separately, and I use a linear approximation architecture. The approximate value function is a linear combination of  $K$  basis functions

$$\tilde{V}(\mathbf{x}, m, \beta) = \sum_{k=1}^K \phi_k(\mathbf{x}) \beta_{km}. \quad (\text{B.12})$$

Here  $\mathbf{x} \equiv (\mathbf{a}, \rho, t)$  is a shorthand notation for the whole vector of state variables excluding the shock realisation,  $m \in \{1, \dots, M\}$  is the index of shock realisation (corresponding to  $s$  which is the actual value of shock).  $\phi_k(\mathbf{x})$ ,  $k = 1, \dots, K$  are basis functions (real functions of the underlying states), and  $\beta = (\beta_{km})_{k=1, \dots, K, m=1, \dots, M}$  is a matrix of real parameters where column  $m$  corresponds to value function for shock realisation with index  $m$ .

In the following discussion I also use the term 'features' to refer to the basis functions. This is a common terminology in machine learning.

Let  $\beta^r$  denote parameters at the beginning of iteration  $r$  of VFI algorithm, with  $\beta^0$  being some initial guess. At iteration  $r$  of value function iteration, we solve the maximisation step in Bellman's equation for  $S$  points in the state space  $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(S)})$  with the shock realisations saved separately as  $(m^{(1)}, \dots, m^{(S)})$ , using  $\tilde{V}(\mathbf{x}^{(s)}, m^{(s)}, \beta^r)$  to obtain continuation values, and save the optimal values as  $V^{[1]}, \dots, V^{(S)}$ . The updated parameter vector  $\beta^{r+1}$  is chosen to minimise  $M$  regularised least squares cost functions

$$J(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(S)}, m^{(1)}, \dots, m^{(S)}, \beta) = \frac{1}{2S} \sum_{m^{(s)}=m} (\tilde{V}(\mathbf{x}^{(s)}, m^{(s)}, \beta) - V^{(s)})^2 + \frac{\lambda}{2S} \sum_{k=2}^K \beta_{km}^2, \quad (\text{B.13})$$

$m = 1, \dots, M$ . The  $M$  regressions (finding columns of  $\beta$  minimising the  $M$  cost functions) are independent and can be performed separately. The parameter  $\lambda \geq 0$  is so called regularisation parameter. The case of  $\lambda = 0$  is standard OLS. Larger values of  $\lambda$  tend to help with the

problem of 'overfitting', i.e., improve approximation accuracy outside of sample<sup>3</sup>. The value of  $\lambda$  is chosen experimentally.

Estimation of parameters can be done directly using well-known linear algebra formulas. This is viable only if the number of features is relatively low (approximately less than 1,000) because otherwise the matrix operations needed take too much time and we may run out of computer memory. To deal with cases where direct solution does not work in practice, I implement a parallel version of a stochastic gradient descent algorithm to estimate the parameters. This algorithm scales extremely well compared to direct linear algebra solution, and it allows us to efficiently estimate the parameters even if the number of features is rather large (I tested the algorithm for 4,000 features).

The  $S$  points in the state space which we use to get an updated vector of coefficients can be chosen in a number of ways. I chose a deterministic grid invariant over the VFI iterations. This is a natural starting point and it is convenient for computation reasons<sup>4</sup>.

The hope is that even though convergence is not guaranteed, in practice we can achieve convergence by choosing appropriate parameters, primarily those governing the set of points in state space used for estimating parameters  $\beta$ , the regularisation parameter  $\lambda$ , and the set of features (basis functions). The value of parameter  $\lambda$  is crucial in this context. By increasing  $\lambda$ , we tend to improve fit outside of the sample (see Nguyen (2012)), and hence reduce the chance of divergence. If we increase  $\lambda$  too much, however, we obtain a bad approximation both within and outside the sample, as the values of all the coefficients except for the constant term are pulled closer to zero.

As discussed in the main text, the fitted value function algorithm did not result in robust convergence. There are several promising approaches which may facilitate the stability of the algorithm, and which are a topic for future work. Firstly, instead of using points in the state space on a fixed grid to estimate the parameters of the approximate value function, we can use simulation to obtain the sample. Bertsekas (2012) shows that not taking into account the likelihood of points in the state space being actually visited when collecting samples for estimating parameters is a potential source of divergence. We can also use some kind of robust regression to estimate the coefficients. Boyan and Moore (1995) show that the issue of divergence can be fixed by a non-expansive projection (which is not satisfied by linear regression). We can also use even more points in the state space to estimate the

---

<sup>3</sup>Note that  $\beta_1$  which is the coefficient associated with the constant term is not regularised, which is a conventional choice common in machine learning applications. If this term were included we would obtain a standard ridge regression estimator used in econometrics to deal with multicollinearity. Haykin (2009) discusses regularisation in more detail.

<sup>4</sup>With a fixed grid, the values of features are iteration-invariant. If the number of features is relatively small (up to around 1,000) then we can use linear algebra methods (singular value decomposition), and pre-compute some matrices involving computationally costly operations at the outset.

parameters, as Munos and Szepesvári (2008) show that this leads to asymptotically lower error bounds. Finally, we could implement a non-linear approximation architecture, as even the best feasible linear approximation may not be good enough to approximate the value function accurately in every iteration. This approach is advocated by Haykin (2009) who suggests the use of neural networks.

## Discrete value function iteration (algorithm)

Assume that we have  $P$  processors available, indexed by  $p = 1, \dots, P$ . I describe the parallelisation part of the implementation at a relatively abstract level, which can be implemented in a number of programming languages<sup>5</sup>.

### 1. Choose value of parameters related to discretization:

- $N_a, N_\rho, N_t$  - number of grid points for each of state variables
- $\underline{a}_1, \bar{a}_1, \underline{a}_2, \bar{a}_2, \underline{\rho}, \bar{\rho}$  - grid boundaries (minimum and maximum) for each of the state variables.
- $\varepsilon > 0$  - stopping rule threshold,  $R_{max}$  - maximum number of iterations.

### 2. Partition grid points between processors. $N = N_a^2 N_\rho N_t M$ is the total number of grid points. Define lower and upper index bounds for all processors $L^{[p]}, U^{[p]}$ , such that the range of indices for all processors combined covers the whole range and there is no overlap, i.e., $\bigcup_p \{L^{[p]}, \dots, U^{[p]}\} = \{1, \dots, N\}$ , $\bigcap_p \{L^{[p]}, \dots, U^{[p]}\} = \emptyset$ . Define invertible correspondence $\Upsilon : \{1, \dots, N\} \mapsto \{1, \dots, N_a\} \times \{1, \dots, N_a\} \times \{1, \dots, N_\rho\} \times \{1, \dots, N_t\} \times \{1, \dots, M\}$ . This correspondence transforming index in the range $1, \dots, N$ into 5 indices for the respective state variables can be defined in a number of ways<sup>6</sup>. The five outputs of $\Upsilon$ are denoted as $\Upsilon_1, \dots, \Upsilon_5$ .

### 3. Construct grids for each of the states and save these in vectors $G_{a_1}, G_{a_2}, G_\rho, G_t, G_s$ , where the 'grid' $G_s$ is just vector of realisations of Markov chain which we already have as part of the problem setting. The first and last points of the grids satisfy $G_{a_1}[1] = \underline{a}_1, G_{a_1}(N_a) = \bar{a}_1, G_{a_2}[1] = \underline{a}_2, G_{a_2}(N_a) = \bar{a}_2, G_\rho[1] = \underline{\rho}, G_\rho(N_\rho) = \bar{\rho}, G_t[1] = 1, G_t(N_t) = T$ . The points in between can be chosen such that the grid is equispaced or

<sup>5</sup>The program was written using Fortran 2008 standard, which is very convenient because it includes direct support of parallelisation via coarrays. See Metcalf et al. (2011) for an introduction.

<sup>6</sup>The best way to define  $\Upsilon$  depends on the programming language and compiler. For example, Fortran language standard prescribes that multi-dimensional arrays are stored in physical memory as an array in which the left-most indices are varied the first, and if  $\Upsilon$  has this property performance is optimised. In other languages such as C the order is compiler-specific.

denser in a particular area where we want the solution to be more accurate. In practice this choice made little difference.

4. Let  $\bar{V}_r^{[p]}$  denote discrete approximation of value function on processor  $p$  at the beginning of iteration  $r$ . This is a five-dimensional array with element  $(i_1, i_2, i_3, i_4, i_5)$  corresponding to value at point in state space  $(G_{a_1}(i_1), G_{a_2}(i_2), G_\rho(i_3), G_t(i_4), G_s(i_5))$ . Initialise (assign starting values to)  $\bar{V}_1^{[1]}$ , i.e., approximation of value on processor 1 at the start of iteration 1. The algorithm is not sensitive to starting values so setting all elements to zero for example works fine.
5. Iterate on value function. For  $r = 1, \dots, R_{max}$ :
  - (i) Send  $\bar{V}_r$  from processor one to all other processors:  $\bar{V}_r^{[p]} = \bar{V}_r^{[1]}, p = 2, \dots, P$ .
  - (ii) At every processor (in parallel), solve for the part of next-iteration value function belonging to the processor. For  $l^{[p]} = L^{[p]}, \dots, U^{[p]}$ :
    - (a) Get current state:  $a_1 = G_{a_1}(\Upsilon_1(l^{[p]}))$ ,  $a_2 = G_{a_2}(\Upsilon_2(l^{[p]}))$ ,  $\rho = G_\rho(\Upsilon_3(l^{[p]}))$ ,  $t = G_t(\Upsilon_4(l^{[p]}))$ ,  $s = G_s(\Upsilon_5(l^{[p]}))$ .
    - (b) For this state, perform the maximisation of the right-hand side of the Bellman equation. Instead of the true value function  $V$  use  $\tilde{V}$ , which is a multi-dimensional interpolation which uses values on the grid  $\bar{V}_r^{[p]}$ . Choice of  $\tilde{V}$  and the maximisation step are discussed later in this appendix. Save the optimal value in element  $l^{[p]}$  of vector  $\hat{V}^{[p]}$  (elements of this vector on processor  $p$  are indexed from  $L^{[p]}$  to  $U^{[p]}$ ).
  - (iii) Wait for all processors to finish step (ii). On processor 1, read the values of  $\hat{V}^{[p]}$  from all processors  $p = 1, \dots, P$ , and save these in vector  $\hat{V}$  indexed from 1 to  $N$ . Now we have the updated value function on processor one as a vector.
  - (iv) On processor 1, reshape the vector  $\hat{V}$  into the five-dimensional array form. Formally, for  $l = 1, \dots, N$ :  $\bar{V}_{r+1}^{[1]}(\Upsilon_1(l), \Upsilon_2(l), \Upsilon_3(l), \Upsilon_4(l), \Upsilon_5(l)) = \hat{V}(l)$ .
  - (v) If a stopping rule is satisfied, exit the loop over  $r$ . The stopping rule used is maximum absolute difference between  $\bar{V}_{r+1}^{[1]}$  and  $\bar{V}_r^{[1]}$  being less than  $\varepsilon$ .
6. Save the latest value  $\bar{V}_r^{[1]}$  obtained by iterating as  $\bar{V}$ . This is the approximate value function solution.

For any initial conditions we can solve the  $t = 0$  period problem using interpolates of  $\bar{V}$  to get continuation value. We then recover the next-period states and policy. In the following periods we do the same thing but we use the recursive formulation for periods  $t \geq 1$  rather

than the first-period problem. The solution of the first period problem and the simulation is very similar to a maximisation step in the value iteration at a point in state space which will be described below.

**Interpolation** In solving the maximisation step at a point in state space, we need to be able to evaluate continuation values  $\tilde{V}(a'_1, a'_2, \rho', t', s)$  for all next-period states, not just those falling onto the grid. We use quadrilinear interpolation, which is a straightforward generalisation of bilinear interpolation<sup>7</sup>. The advantage of this method is that it is very fast, evaluated by a single formula (once we find the closest two points on the grid for every dimension) without any calls to other functions or subroutines. This matters because continuation value is computed very often, as the maximisation step at every point in state space relies on an iterative algorithm. Another advantage of this method is stability. In Chapter 4, various interpolation techniques were implemented, and Appendix C4 contains a discussion comparing their merits.

**Choice of initial guess - CTS (course to smooth) algorithm** The value iteration seems to almost always converge to a similar solution irrespective of initial guess, so we can use a naive guess such as  $\bar{V}_1 \equiv 0$ . If we choose a good initial guess, we get a more accurate solution in a given time. We implement Course to Smooth (CTS) algorithm to obtain a good guess. The basic idea is to perform the value function iteration for a low number of grid points, which is much faster, then increase the number of grid points, and use the value function obtained on the coarse grid as initial guess for the value function on the smooth grid (using the same interpolation method as elsewhere). The step from coarse to smooth grid can be repeated several times, and the number of grid points per variable is roughly doubled at every step. Other authors have also used a solution obtained for a lower number of grid points as an initial guess in dynamic programming, although implementation details are different. See for example Aruoba et al. (2006) or Heer and Maussner (2011).

## Other acceleration techniques

In the discrete value iteration, I use Howard's acceleration algorithm, also known as optimistic policy iteration. The policy function changes very little in between iterations of VFI algorithm, at least after a few dozen of initial iterations, and computation of policy function is where the algorithm spends the most time. Convergence can thus be accelerated if we skip (possibly

<sup>7</sup>See for example Judd (1998) for exposition of bilinear interpolation algorithm.

dozens of times) the optimisation step and use a policy function obtained from a previous iteration.

The parallel discrete value function iterations spends a relatively great amount of time by sending data between processors, because every iteration, each processor obtains a copy of the whole updated value function. It might be possible to achieve further performance gains by skipping the synchronisation step in some iterations, i.e., by updating the value function locally only, and synchronising every given multiple of iteration number. Bertsekas (2012) discusses this so called asynchronous value iteration.

### Fitted value function iteration (algorithm)

I describe a version of the algorithm which draws the points in state space used for estimation randomly. In the version of the algorithm outlined here different points in state space are chosen at every iteration of value iteration. The alternative would be drawing the points only once at the beginning of the first iteration. The only real advantage of drawing the points once only would be in the case of relatively few points (order of hundreds of thousands) and features (less than 1,000), in which case direct solution methods are feasible for estimating the parameters, and we can pre-compute some matrices at the outset to save time. In practice more points in state space and features are necessary to accurately approximate the value function. If we draw points repeatedly we can use optimal policy (at current iteration) to simulate samples, which may enhance convergence properties as discussed by Bertsekas (2012).

Assume that we have chosen features  $\phi_1, \dots, \phi_k$  in the linear approximation of value function  $\tilde{V}(\mathbf{x}, m, \beta) = \sum_{k=1}^K \phi_k(\mathbf{x}) \beta_{km}$ .

#### 1. Choose value of parameters:

- $N_{st}, N_{sim}$  - number of points in state space used for estimating the parameters.  $N_{st}$  is the number of chosen initial points in state space on every processor,  $N_{sim}$  is the length of simulation starting at each of these points. Total number of points is  $PN_{st}N_{sim}$ .
- Parameters governing how the initial points are chosen. These can for example be parameters of truncated multivariate normal distribution, or multivariate uniform distribution.
- $\varepsilon > 0$  - stopping rule threshold,  $R_{max}$  - maximum number of iterations.

#### 2. Let $\beta_r^{[p]}$ denote value of parameters on processor $p$ at the beginning of iteration $r$ . Initialise (assign starting values to) $\beta_1^{[1]}$ , the parameters on processor 1 at the start of



iteration 1. Unlike in the case of discrete approximation of value, the solution depends on the initial guess.

3. Iterate on value function. For  $r = 1, \dots, R_{max}$ :

- (i) Send parameters from processor one to all other processors:  $\beta_r^{[p]} = \beta_r^{[1]}, p = 2, \dots, P$ .
- (ii) At every processor (in parallel), draw  $N_{st}$  initial points in state space  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N_{st})}$ , with corresponding shock realisation indices  $m^{(1)}, \dots, m^{(N_{st})}$ . The distribution used to draw the points is discussed later. These points are different at every processor but the superscript  $[p]$  is omitted because we never access these samples from other processors directly (we only access statistics).
- (iii) At every processor generate  $N_{st}$  samples of length  $N_{sim}$ , starting at the initial points drawn previously. The policy that we use to simulate is the choice maximising expected current plus continuation value using the parametric approximation  $\tilde{V}(\mathbf{x}, m, \beta_r)$  in place of the continuation value. Collect all the samples at a processor into vectors of states  $(\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(S)})$ , and the corresponding shock realisations kept separately as  $(m^{(1)}, \dots, m^{(S)})$ , where  $S = N_{st}N_{sim}$ . The optimal values starting at these states are saved as  $(V^{(1)}, \dots, V^{(S)})$ .
- (iv) Use a parallel implementation of gradient descent to estimate parameters by minimising regularised least squares cost function, following Nguyen (2012). This is an iterative algorithm, and in every iteration we require the gradient of the cost function over the whole sample on all processors. This can be obtained as summation of one term per processor which involves sample on that processor only. The updating is performed on processor 1. At the end of the iterative algorithm we get the new value of parameters on processor 1:  $\beta_{r+1}^{[1]}$ .
- (v) We can check a stopping criterion in terms of change of value of parameters. In the case of highly correlated features (which is the case if we use high degrees of polynomials) the coefficients may fluctuate quite a lot unless the value of regularisation parameter is relatively high.

4. Save the latest value  $\beta_r^{[1]}$  obtained above as  $\beta$ . This is the approximate value function solution.

**Distribution of initial points in state space** In every iteration, we draw  $N_{st}$  points in the state space from some multivariate distribution, and then simulate for further  $(N_{sim} - 1)$  periods. We can select the initial points in many ways and it is not obvious which methods

is best. For example, we can use a multivariate normal distribution (truncated to get rid of influence of outliers), a multivariate uniform distribution, or even a deterministic set of points on a grid.

**More details on gradient descent implementation** The gradient descent estimation of parameters can be accelerated by not using the whole sample in every, but rather a small set of randomly selected observations in the training set (for example 10 per processor rather than thousands). This is known as mini-batch gradient descent and works well in practice (see Nguyen (2012)).

The main advantage of gradient descent is that it allows us to solve for parameters even in a case where the number of grid points and/or the number of features is too high for direct solution methods to be feasible. It is also a learning algorithm in the sense that it uses an initial guess of parameters and iteratively improves on this guess. It is thus particularly well suited to value iteration where the value function should differ less and less in later iterations. The parallelisation is efficient because each processor sends only a  $(K + 1)$ -dimensional vector to processor 1, where  $K$  is the number of features, rather than a vector involving the number of observations at every processor, which would be a much larger object in practice.

In principle, we could use every point in the state space only once and update the relevant summation expressions 'on the fly' (we would not use the updated coefficients until a given number of iterations or a convergence criterion were satisfied). This would essentially be a mini-batch gradient descent. The advantage of this algorithm would be irrelevance of memory limits. In practice, this would be inefficient because obtaining a value at a point in state space is very computationally costly, and it would be wasteful to use this information in an iterative algorithm for estimating coefficients only once. Furthermore, if we keep all the points collected, we can use robust estimation techniques such as discarding outliers or least median squares. The memory limitation imposed by this choice are not relevant conditional on an appropriate hardware architecture (many CPUs with separate operating memory).

**Choice of basis functions** The convergence of the algorithm is conditional on a good choice of basis function (features). After experimenting with multiple options I settled on tensor products of Chebyshev polynomials in the underlying state variables. The order of polynomials is different for different variables. In particular, it appears that including higher order than 5 polynomials of  $\rho$  does not really improve the approximation, which is not the case for asset holdings  $(a_1, a_2)$ .

## Solving the maximisation step

Here I address the problem of solving the maximisation step at a point in state space, assuming that have some approximation of value function  $\tilde{V}$ , which can be either a discrete or parametric approximation. The discussion is done in the context of the model of Chapter 2.

We maximise the right-hand-side of the Bellman equation subject to all the constraints. The strategy is to express most choice variables from the constraints, and plug these into the other constraints and the objective function, rather than solving a higher-dimensional problem with the original constraints directly<sup>8</sup>. Recall that  $M$  is the number of realisations of government expenditure shock process (Markov chain). We get an unconstrained optimisation problem, where the choice variables are  $\{c_1(s)\}_{s=s_1, \dots, s_M}, \{c_2(s)\}_{s=s_1, \dots, s_{M-1}}$ , so that we maximise over state-contingent consumption plan excluding consumption of agent 2 in the last state<sup>9</sup>. For two agents and  $M = 1$  (deterministic case), we get a one-dimensional maximisation problem.

Constraints (A.15) - (A.18) are used to recover current productivities of all agents, and current government expenditure. These are essentially exogenous variables. Constraints (A.13) and (A.14) are used to express labour supplies  $[l_1(s), l_2(s)]_s$  as function of state-contingent consumption plan. Constraints (A.10) and (A.11) are used to express  $[a'_1(s), a'_2(s)]_s$  as functions of state-contingent consumption and labour plans. Equation (A.12) is used to express  $c_2(M)$  as function of  $\{c_1(s)\}_{s=s_1, \dots, s_M}, \{c_2(s)\}_{s=s_1, \dots, s_{M-1}}$ .

For the special case of two agents and the common utility function given by used in the main text (log in consumption, CRRA in labour), the labour supplies can be expressed as

$$l_1(s) = \frac{\pi_1 c_1(s) + \pi_2 c_2(s) + g}{\pi_1 \theta_1 + \pi_2 \theta_2 \left( \frac{\theta_2 c_1(s)}{\theta_1 c_2(s)} \right)^{1/\gamma}} \quad (\text{B.14})$$

$$l_2(s) = \frac{\pi_1 c_1(s) + \pi_2 c_2(s) + g}{\pi_1 \theta_1 \left( \frac{\theta_1 c_2(s)}{\theta_2 c_1(s)} \right)^{1/\gamma} + \pi_2 \theta_2} \quad (\text{B.15})$$

<sup>8</sup>This approach is practical for easier debugging, and handling of issues of non-convex feasible sets. At a point in state space, due to non-convexity of feasible sets it is important for stability that we start a maximisation algorithm at a good guess, as global maximisation algorithms that initialise at multiple points in the feasible set are slow. We thus need to store optimal solutions from previous iterations (the optimal policy changes slowly over iterations so this is a good guess). Storing a high-dimensional vector of choice variables would become infeasible. This advantage is relevant to the discrete approximation case and to the fitted value iteration if sample of points in state space is generated once only.

<sup>9</sup>This choice is arbitrary, we can express any single element of the state-contingent consumption plan as a function of the other elements.

for all  $s$ . Next-period marginal-utility-adjusted asset holdings can be expressed as ( $\forall s$ )

$$a'_1(s) = \frac{1}{\beta} \left\{ \frac{a_1}{c_1(s) \mathbb{E}_{s_-} \left[ \frac{1}{c_1} \right]} + Bl_1(s)^{1+\gamma} - A \right\} \quad (\text{B.16})$$

$$a'_2(s) = \frac{1}{\beta} \left\{ \frac{a_2}{c_2(s) \mathbb{E}_{s_-} \left[ \frac{1}{c_2} \right]} + Bl_2(s)^{1+\gamma} - A \right\}. \quad (\text{B.17})$$

From equation (2.20), the second agent's last-shock-realisation consumption can be expressed as

$$c_2(s_M) = \rho p_M \left[ \sum_{i=1}^M \frac{p_i}{c_1(i)} - \sum_{i=1}^{M-1} \frac{\rho p_i}{c_2(i)} \right]^{-1}, \quad (\text{B.18})$$

where  $p_i$  is the probability that current-period shock realisation will be  $s_i$ ,  $i = 1, \dots, M$ . These probabilities depend on state  $s_-$  (they form a row vector of the transition matrix of the shock process). I successively plug these expressions into all other constraints and the objective function. What we are left with is an objective function of  $[c_1(s)]_s, [c_2(s)]_{s \in \{s_1, \dots, s_{M-1}\}}$  only.

**Choice of boundaries of state space and loss function approach** Given a choice of boundaries of state space, it may happen that at some points (usually in some 'corner') there is no choice of consumption for which all next-period states lie within the boundaries, or at least it is not easy to find numerically. Then we can not use interpolation to obtain continuation value, and using extrapolation is problematic for stability. Choosing arbitrarily wide state space does not resolve this problem.

A loss function approach is adopted, in which a choice implying a next-period state outside the grid is permitted, but a convex loss function penalty is added to such a choice. The continuation value for choices outside of the grid are obtained by nearest neighbour extrapolation, with the added loss. The share of points at which the choice is outside of grid is very small (for the results reported here) and the violations are quantitatively small. Because of the convex loss function, even if the initial guess implies a choice outside of grid boundaries, in following iterations the choice usually converges to a point in which the violation of grid boundaries is essentially zero. The grid boundaries are chosen experimentally, so that their change does not quantitatively significantly change the solution.

# Appendix C

## Chapter 4 appendices

### Appendix C1: Alternative specifications

This appendix contains definitions and sequential formulations of the Ramsey problem for the two alternative specifications: the full union, and autarky.

#### Full union

**Definition** (*allocation, asset profile, price system, tax policy*)

An *allocation* is a sequence  $\{c_t^i, l_t^i\}_{i,t}$  of consumption and labour supply in the two countries. An *asset profile* is a sequence of households' assets and pooled government debt  $\{\{b_{i,t}\}_{i,t}, \{B_t\}_t\}$ . A *price system* is a sequence  $\{R_t\}_t$  of gross interest rates on the one-period bonds. A *tax policy* is a sequence of tax rates on labour income in both countries  $\{\tau_t^i\}_t$ .

**Definition** (*competitive equilibrium*)

For a given initial asset profile  $\{\{b_{i,-1}\}_i, \{B_{-1}\}\}$ , and a given tax policy  $\{\tau_t^i\}_{i,t}$ , a *competitive equilibrium* is an allocation  $\{c_t^i, l_t^i\}_{i,t}$ , an asset profile  $\{\{b_{i,t}\}_{i,t}, \{B_t\}_t\}$  and a price system  $\{R_t\}_t$ , such that:

1. The allocation  $\{c_t^i, l_t^i\}_{i,t}$  and households' assets  $\{b_{i,t}\}_{i,t}$ , maximise the utility (4.1) of household in each country, subject to the respective budget constraint (4.14).
2. The resource constraint (4.3) is satisfied.
3. The pooled governments' budget constraints (4.12) are satisfied.
4. Market-clearing in the single asset market (4.15) is satisfied.
5. The asset profile is bounded.

**Definition** (*socially optimal competitive equilibrium - Ramsey plan*)

Given an initial asset profile  $\{\{b_{i,-1}\}_i, \{B_{-1}\}\}$ , a *socially optimal competitive equilibrium* (Ramsey plan) is a tax policy  $\{\tau_t^{i*}\}_{i,t}$ , an allocation  $\{c_t^{i*}, l_t^{i*}\}_{i,t}$ , an asset profile  $\{\{b_{i,t}^*\}_{i,t}, \{B_t^*\}_t\}$ , and a price system  $\{R_t^*\}_t$  such that

1. Given the initial asset profile  $\{\{b_{i,-1}\}_i, \{B_{-1}\}\}$  and the tax policy  $\{\tau_t^{i*}\}_{i,t}$ , the other elements of the socially optimal competitive equilibrium constitute a competitive equilibrium.
2. Given the initial asset profile  $\{\{b_{i,-1}\}_i, \{B_{-1}\}\}$ , there is no other feasible tax policy  $\{\tau_t^i\}_t \neq \{\tau_t^{i*}\}_t$ , for which a competitive equilibrium exists in which a strictly greater value of the social welfare function (4.2) is attained.

Following similar steps as in the baseline model we arrive at the following Ramsey problem<sup>1</sup>

$$\max_{\{c_t^i(s^t), l_t^i(s^t), b_{i,t}(s^t)\}_{t=0, \forall i, s^t}} \mathbb{E}_0 \sum_i \alpha_i \pi_i \sum_{t=0}^{\infty} \beta^t u(c_t^i, l_t^i)$$

subject to

$$b_i(s^t) = b_{i,t-1}(s^{t-1}) \frac{u_{c,t-1}^i(s^{t-1})}{\beta \mathbb{E}_{t-1}(u_{c,t}^i)} - \frac{u_{l,t}^i(s^t)}{u_{c,t}^i(s^t)} l_t^i - c_t^i \quad (\text{C.1})$$

$$0 = \sum_i \pi_i (c_t^i(s^t) + g_t^i(s^t) - \theta_i l_t^i(s^t)), \forall t, s^t \quad (\text{C.2})$$

$$\frac{u_{c,t}^1(s^t)}{\mathbb{E}_t(u_{c,t+1}^1)} = \frac{u_{c,t}^2(s^t)}{\mathbb{E}_t(u_{c,t+1}^2)}, \forall t, s^t \quad (\text{C.3})$$

Once we have solved for the allocation and households' assets, the government debt can be obtained using the market-clearing condition. Tax policy and the common interest rate is obtained using the first-order conditions of the households' problems as usual. A recursive formulation of the problem is similar to the problem in Chapter 2, with the difference that labour productivity is constant (so there is one fewer state variable), and there is no constraint following from equality of tax rate across households (countries). The model is equivalent

<sup>1</sup>Because there is now a single consolidated government constraint, it is more convenient to choose the constraint of the government as the one which we omit, rather than one of the constraints of the households. The formulation is similar to the problem in Chapter 1. From a mathematical point of view, the case of full union is essentially equivalent to the case of a single country with expenditures  $g_t = \sum_i \pi_i g_t^i$ , in which the tax rate can vary between households (so in the case of two countries there is one more degree of freedom per shock realisation).

to a closed-economy model with heterogeneous agents (as in the previous two chapters), expenditure process  $g_t = g_t^1 + g_t^2$ , and agent-specific tax rate on labour.

## Autarky

In autarky the cross-country borrowing is ruled out, so  $b_j^i = 0$  for all  $i \neq j$ . The definitions of equilibrium, Ramsey plan, and the sequential formulation of the Ramsey problem in the model of partial union still apply, with the addition of these constraints, and the removal of the no-arbitrage constraint.

The countries are effectively isolated, so we can solve a one-country optimal policy problem for each country separately. The one-country model is the model of Aiyagari et al. (2002) which is comparatively easy to solve because we need to keep track of only three state variables ( $b_i^i, u_i^c, s_-$ ).

## Appendix C2: Further results

Figures C.1 and C.2 show the optimal asset profile chosen in the partial union in response to purely idiosyncratic shocks (discussed in Section 4.4), for a particular realisation of government expenditures, and two different normalisations. Initially, the shock realisations alternate, and the net external debt stays close to zero. Country 1 was more 'unlucky' in later periods since it experienced prolonged periods with government expenditure greater than in country 2. As the net external debt of a country increases in periods of high expenditures, the net external debt of country 1 gradually increased up to around 40% of GDP. The (net) debt of each government stays close to zero, and the households do most of the borrowing between each other.

As was discussed in the main text (Result 1), only the net external debt is determined in equilibrium, and we can normalise for example by choosing  $b_1^2 = 0$  (so that households of country 1 do not buy bonds issued by government of country 2), or  $b_2^1 = 0$ . for both normalisations, the assets used as states ( $B^1, B^2, b_1^{ex}$ ) are the same, as are the net assets of each household (which appear in their respective budget constraints). In the first case (normalisation  $b_1^2 = 0$ , Figure C.1), the households of country 1 are borrowing from their government, which is in turn borrowing from the households of country 2. In the second case (normalisation  $b_2^1 = 0$ , Figure C.2), the households in country 1 are borrowing from the government of country 2, which is borrowing from its own households.

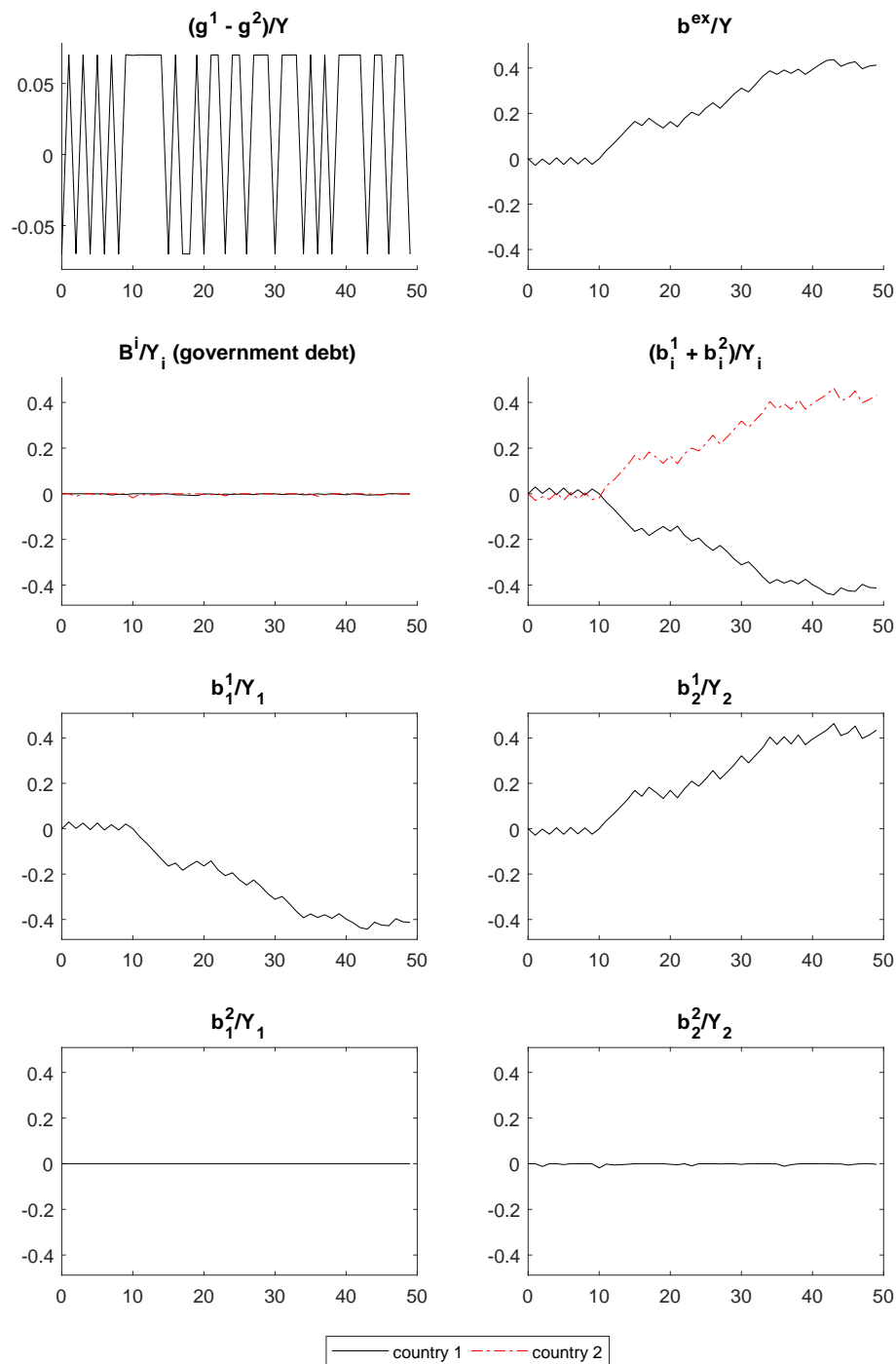
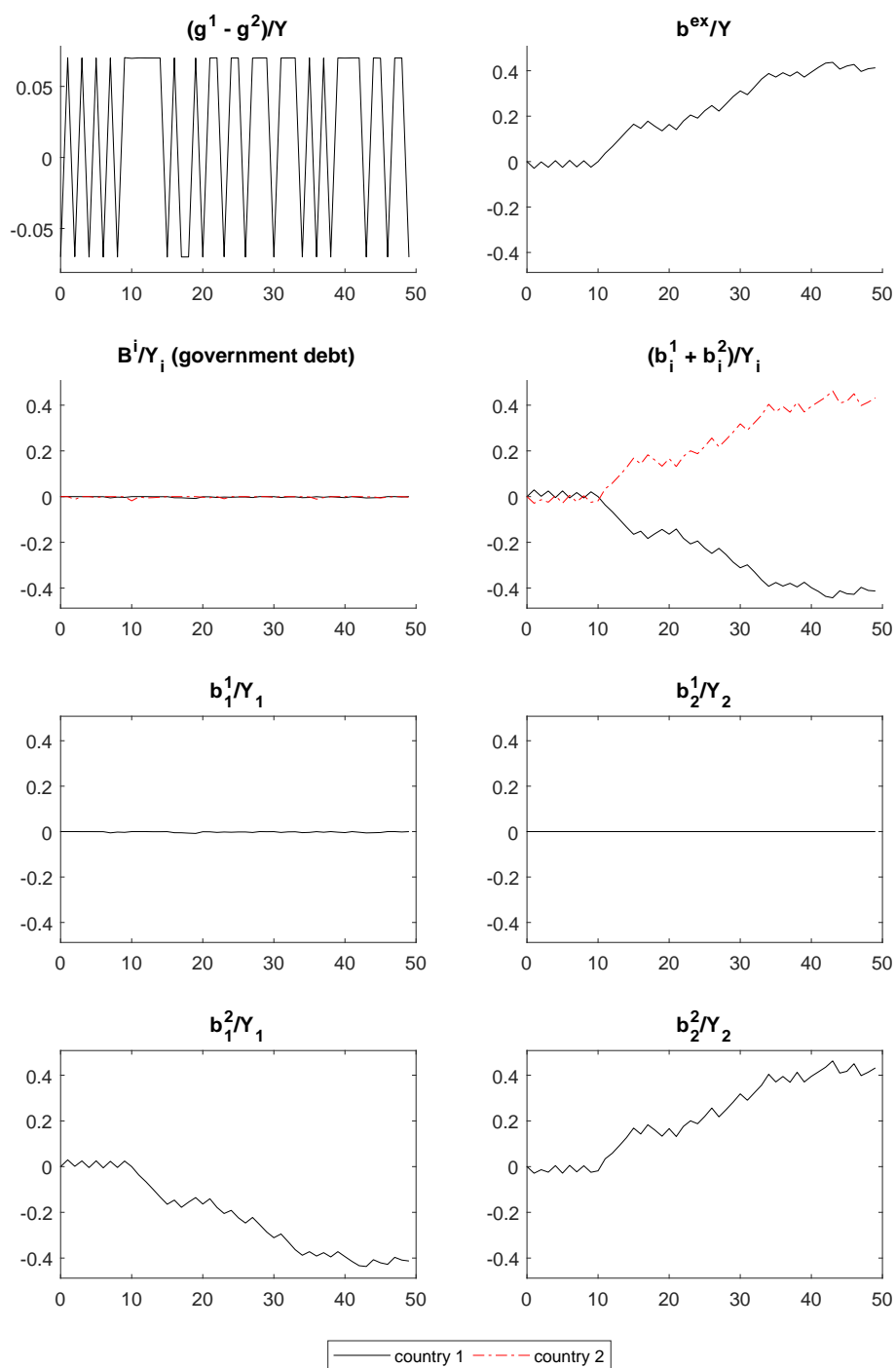
Fig. C.1 Optimal portfolio with normalisation  $b_1^2 = 0$ 



Fig. C.2 Optimal portfolio with normalisation  $b_2^1 = 0$ 

## Appendix C3: Calibration

Table C.1 summarises the calibration targets and parameters. We solve the model for different values of coefficient of risk aversion  $\sigma$ . For every value of  $\sigma$  we choose the value of parameter  $B$  so that the steady-state labour supply is kept the same for all  $\sigma$ . If we did not make this adjustment then the government expenditure shocks would be either a greater or a smaller share of steady-state consumption (and steady-state tax rate would deviate from the target), and this would invalidate welfare comparisons. Table C.2 lists the pairs  $(\sigma, B)$  which we use.

Table C.1 Summary of calibration

Calibration targets	
annual interest rate	5.26%
Frisch elasticity of labour	0.5
labour supply	0.7
tax rate	0.25
Parameters	
$I$ (number of countries)	2
$(\pi_1, \pi_2)$ (mass of households)	(0.5, 0.5)
$\beta$ (discount factor)	0.95003
$\gamma$ (utility function parameter)	2
$(\bar{l}_1, \bar{l}_2)$ (maximum labour)	(3, 3)
$(\theta_1, \theta_2)$ (labour productivities)	(1, 1)
$(g_L, g_H)$ (expenditure shock realisations)	$(0.93\bar{g}, 1.07\bar{g})$
$P_{1,1}$	0.5
$P_{2,2}$	0.5
$\bar{g}$ (average gov. expenditure)	0.175
$(b_1^{ex}, B^1, B^2)$ (initial assets)	(0, 0, 0)

Table C.2 Utility function parameters

Pairs of utility parameters			
$\sigma$	1	2	3
$B$	2.9155	5.5532	10.5776

## Appendix C4: Numerical appendix

This appendix discusses solution of the baseline model (partial union). We solve the model using value function iteration (VFI) on a discrete grid as in the previous two chapters. In this

appendix we address only the most important features of the solution algorithm which are different. We use all of the acceleration techniques which are discussed in the context of the discrete approximation VFI in Appendix B3 (parallelisation, CTS algorithm, Howard's algorithm).

Firstly, we give a recursive formulation which we use, and we describe the implications of the differences relative to the previous two chapters for the solution algorithm. We then address the maximisation step at a point in state space in the VFI, and the topics of exploration and choice of interpolation scheme, which are both related to the non-convexities inherent to this problem<sup>2</sup>.

### Recursive formulation and the maximisation step in VFI

We can reduce the number of state variables by one (relative to the formulation of Section 4.3.2) using the following recursive formulation, which utilises the fact that the bonds and marginal utilities only enter the constraints as a product of the two, and as a ratio of the marginal utilities. The new state variables are

- $\mathbf{a} = (a^1, a^2, a_1^{ex}) \equiv (\frac{1}{\beta}B^1u_c^1, \frac{1}{\beta}B^2u_c^2, \frac{1}{\beta}b_1^{ex}u_c^1)$ : last-period MU-adjusted assets (total debt of country 1, of country 2, and net foreign debt of country 1)
- $\rho = \frac{u_c^1}{u_c^2}$ : last-period ratio of marginal utilities of consumption of HHs in country 1 and 2. This is where we save one state variable, we only need to keep track of the ratio as opposed to both levels of marginal utilities.
- $s_-$ : last-period shock realisation (for i.i.d. shocks this disappears)

The Bellman equation for period  $t \geq 1$  becomes

$$V(\mathbf{a}, \rho, s_-) = \max_{[c^i(s), l^i(s), a^i(s)']_{i,s}, [a_1^{ex}(s)']_s} \sum_s Pr(s|s_-) \left[ \sum_{i=1}^2 \alpha_i \pi_i u(c^i(s), l^i(s)) + \beta V(\mathbf{a}(s)', \frac{u_c^1(s)}{u_c^2(s)}, s) \right]$$

<sup>2</sup>The problem was already present in the models solved in the first two chapters. However, it is much easier to deal with because there are fewer free choice variables. Particularly in the deterministic calibrations (which result in one free variable only) the non-convexities are not a serious problem in practice.

subject to

$$\beta \frac{a^i(s)'}{u_c^i(s)} = \frac{a^i}{\sum_s Pr(s|s_-)u_c^i(s)} + \pi_i g^i(s) - \pi_i \left(1 + \frac{u_l^i(s)}{\theta_i u_c^i(s)}\right) \theta_i l^i(s), \forall i, s \quad (C.4)$$

$$0 = \sum_i \pi_i (c^i(s) + g^i(s) - \theta_i l^i(s)), \forall s \quad (C.5)$$

$$\beta \frac{a_1^{ex}(s)'}{u_c^1(s)} = \frac{a_1^{ex}}{\sum_s Pr(s|s_-)u_c^1(s)} + \pi_1 [g^1(s) + c^1(s) - \theta_1 l^1(s)], \forall s \quad (C.6)$$

$$\rho = \frac{\sum_s Pr(s|s_-)u_c^1(s)}{\sum_s Pr(s|s_-)u_c^2(s)} \quad (C.7)$$

The period 0 problem is almost identical to the original formulation, the only difference is that we replace the value function in the objective function by  $V(\frac{1}{\beta} B_0^1 u_{c,0}^1, \frac{1}{\beta} B_0^2 u_{c,0}^2, \frac{1}{\beta} b_{1,0}^{ex} u_{c,0}^1, s_0)$ .

For the CRRA utility function which is used in the numerical examples, the constraints in the Bellman equation for periods  $t \geq 1$  become

$$a^i(s)' = \frac{1}{\beta c^i(s)^\sigma} \left[ \frac{a^i}{\sum_s Pr(s|s_-)c^i(s)^{-\sigma}} + \pi_i g^i(s) - \pi_i (\theta_i l^i(s) - B c^i(s)^\sigma l^i(s)^{1+\gamma}) \right], \forall i, s \quad (C.8)$$

$$0 = \sum_i \pi_i (c^i(s) + g^i(s) - \theta_i l^i(s)), \forall s \quad (C.9)$$

$$a_1^{ex}(s)' = \frac{1}{\beta c^1(s)^\sigma} \left[ \frac{a_1^{ex}}{\sum_s Pr(s|s_-)c^1(s)^{-\sigma}} + \pi_1 [g^1(s) + c^1(s) - \theta_1 l^1(s)] \right], \forall s \quad (C.10)$$

$$\rho = \frac{\sum_s Pr(s|s_-)c^1(s)^{-\sigma}}{\sum_s Pr(s|s_-)c^2(s)^{-\sigma}}. \quad (C.11)$$

We now address the maximisation step in VFI. This discussion assumes that we have an approximation of the value function  $V$  either from a previous iteration, or an initial guess, and we can thus compute the continuation value for any value of next-period states using interpolation. The choice of interpolation scheme is discussed later in this appendix.

To perform the maximisation step in value function iteration at every point in the grid (given the state), we need to be able to compute the current return plus continuation value. The strategy is to use some constraints to reduce the number of free variables as far as possible. Our choice variables are going to be  $\mathbf{x} \equiv (\{l^1(s)\}_{s=1}^M, \{c^1(s)\}_{s=1}^M, \{c^2(s)\}_{s=1}^{M-1})$ , i.e., consumption and leisure of country 1 for every possible shock realisation, and consumption of country 2 in every shock realisation except the one with the highest index (it is just a convention, we could omit the consumption of any country in any state rather than the consumption of country 2 in this particular state).

Given these choices, we can obtain the missing element of the consumption matrix,  $c^2(M)$ , using constraint (C.11). If  $p_i$  denotes the transition probability conditional on last-period shock realisation  $s_-$  we get the consumption of country 2 in state  $M$  as<sup>3</sup>

$$c^2(M) = (\rho p_M)^{1/\sigma} \left[ \sum_{s=1}^M p_s c^1(s)^{-\sigma} - \rho \sum_{s=1}^{M-1} p_s c^2(s)^{-\sigma} \right]^{-1/\sigma}. \quad (\text{C.12})$$

Then we can use the resource constraint to back out leisure of country 2 for every possible shock realisation. Once we have obtained the full allocation (consumption and labour plan), we can use constraints (C.8) and (C.10) to back out the marginal-utility adjusted asset positions for both countries and all shock realisation  $a^i(s)', a_1^{ex}(s)'$ . We can then compute the continuation value, and adding this to the (expected) current-period return, we can compare values for alternative choices of  $\mathbf{x}$ , looking for one which yields the maximum return.

## Non-convexities and local optima

The feasible set at a point in state space is not convex. The practical manifestation of this problem is that when we perform the maximisation step at a point in state space (using a local numerical optimisation technique), the solution may converge to a point which is only locally optimal, but which is close to the initial guess. At every iteration in VFI algorithm we use the optimal choice from the previous iteration as the initial guess, and the solution may be 'stuck' in a suboptimal local optimum for many iterations. This can in turn affect the value function in the sense that in some regions of state space the approximated value is lower than the true (unknown) value, simply because we are stuck in a bad local optimum, and this may be the case even if conventional convergence criteria for the policy function and the value function are satisfied.

To overcome this problem, it is important to start with a guess of policy function (which is simply a list of choices  $\mathbf{x}$  per each grid point) which is not too far from the optimal choice in the first iteration of VFI<sup>4</sup>. In subsequent iterations, we also need to engage in *exploration*,

<sup>3</sup>If the term in the brackets is less than or equal to zero, then  $c^2(M)$  is either undefined (if  $\sigma \neq 1$ ), or negative (if  $\sigma = 1$ ). In either case this means that we chose a wrong candidate solution. For any  $\rho$ , there are values of the choice vector  $\mathbf{x}$  so that  $c^2(M)$  is defined and positive. For example, we can choose  $\mathbf{x}$  in which the consumption of country 1 is constant across states, and the consumption of country 2 in states  $1, \dots, M-1$  is chosen so that in states  $1, \dots, M-1$  the ratio of marginal utilities is equal to the state  $\rho$ .

<sup>4</sup>To obtain the initial guess at each point in state space before the first iteration of VFI, we use choices  $\mathbf{x}$  which are associated with an autarky steady-state with no initial assets for each country. In the vast majority of points on the grid, this choice is feasible. For some points it is not, in which case we attempt to improve on this guess (by minimising a quadratic loss function which penalises the violation of each constraint). This is similar to what was done in the first two papers and it is discussed in Appendix B3. For the points in state space where asset holdings are low and the state  $\rho$  is relatively close to one, these choices should not be far from optimum.

that is looking for optimal choices relatively far from the initial guess (in addition to the local optimisation).

The fundamental problem associated with exploration is that we need the number of grid points to be great (so that interpolation approximates the value function accurately), and we thus cannot spend a sufficient amount of computation time at every grid point to guarantee that we found a choice very close to a global optimum. The choice of an appropriate exploration approach is an experimental problem. Conceptually it lies in choosing *when* we explore, and *how* we explore.

The question of *when* we explore is equivalent to choosing a probability for every point in the state space that exploration will be done at that point in a given iteration. Our approach to this problem relies on dividing the state space into two regions<sup>5</sup>. Region *A* is a subset of the state space which covers a region where the states (in simulation) spend the vast majority of time, and region *B* is the part of the state space covered by the grid which is not in region *A*. Then we choose a probability of exploration for region *A* and *B* separately (we chose 1 and 0.1). The idea is that while we cannot guarantee that the policy function will be close to a global optimum (and  $V$  will be close to the true value function) everywhere, we make sure that is the case at least in the part of the state space which is visited most frequently<sup>6</sup>.

One further complication is that we need to perform exploration sufficiently frequently. Even if we were able to guarantee that we found the true global optimum at every point in state space, then the next iteration (and particularly after several further iterations) the initial guesses may again be relatively far from the optimal choice. This is particularly a problem if we are using Howard's algorithm (so that maximisation steps are skipped). That is why we set the probability of exploration in region *A* to 1.

As for the question of *how* to explore, after experimentation with several options we settled on using particle swarm optimisation (PSO). This is a heuristic algorithm for global optimisation, which is well suited for the purpose of breaking away from the pull of a local optimum. Poli et al. (2007) discuss the algorithm and survey some applications.

In practice, while exploration is very important particularly from the perspective of obtaining a good approximation of the policy function, the value function tends not to be affected as severely by the problem of local optima. This is the case because alternative local optima tend to be very similar in terms of return, at least in the region of the state space close

---

<sup>5</sup>A description of all details of the algorithm would take too much space. We refer the reader to the description in the sample parameter file *parameters/baseline.txt* which is available online at the address given in Chapter 1, or by request.

<sup>6</sup>One must be careful not to choose the region *A* too small, because otherwise the points in the region are favoured (in terms of value) by virtue of their policy being closer to a globally optimal choice, and the solution will be biased to spend more time in region *A*. In the numerical results reported we chose *A* to cover 1/2 of the grid in every dimension. Region *A* thus covers 1/16 of the grid.

to the initial conditions. However, in some parts of the state space the exploration is very important for preventing divergence (in simulations) which can easily happen if the state gets into a poorly explored region.

The numerical results which are reported in the main text were obtained using value functions which converged<sup>7</sup> even when exploration by particle swarm optimisation is used every iteration (in region A).

## Interpolation scheme

The choice of interpolation scheme is crucially important in numerical dynamic programming. See for example Judd (1998), Heer and Maussner (2009), or Cai and Judd (2014) for a discussion. Ideally, we want to use an interpolation scheme with shape-preserving properties. The advantage is two-fold. Firstly, for a given number of grid points, the accuracy of the solution tends to be more accurate. Secondly, the maximisation step in VFI at a grid point will usually need to make fewer evaluations of the continuation value, because local methods perform better if derivatives of the value function are continuous.

We experimented with several interpolation schemes: quadrilinear interpolation as in the previous two chapters, B-splines (see Miranda and Fackler (2004)), and Shepard's interpolation.

The quadrilinear interpolation (which was used in the first two chapters and is discussed in Appendix B3) is very simple to implement, the evaluation is fast, and it is stable in the sense that the algorithm almost always converges. However, the derivatives of the value function are discontinuous, which often causes local optimisation techniques to require many evaluations to converge, and this somewhat negates the advantage that a single evaluation is very fast.

We solved the model with linear B-splines and the results were very similar to the quadrilinear case but the solution took more time to compute. When we attempted to solve the model with higher-order B-splines (quadratic, cubic), the program usually diverged. This could potentially be resolved by a combination of a different choice of grid than equispaced, more exploration (as discussed above), and a different order for splines for different state variables. Finding a combination of these parameters for which the program robustly converges is a topic for future research.

---

<sup>7</sup>The convergence is in the sense of average change in value and policy function across all grid points. With dozens of millions of grid points it is not feasible to obtain full convergence at every single grid point, particularly with respect to exploration. The simulated series are almost unchanged if we continue iterating on the value function.

We settled on using modified Shepard's interpolation (Shepard (1968)), which yields the best results in terms of stability and accuracy. In Shepard's interpolation the interpolated value is a weighted average of values, where the weight on each point is inversely related to the Euclidean distance of the point from the interpolate. Formally, if  $\mathbf{x}$  is the state at which we want to obtain the interpolated value, and we have pairs  $(v_i, \mathbf{x}_i)$  of values and states, then the interpolated value is

$$V^{shep}(\mathbf{x}) = \sum_{i=1}^N \frac{v_i}{\|\mathbf{x} - \mathbf{x}_i\|^2}. \quad (\text{C.13})$$

Schreiber (2016) proposes the use of Shepard's interpolation for approximating the value function and proves convergence results. The advantage relative to quadrilinear interpolation is that the first and second derivatives are continuous, and local optimisation techniques thus perform better. It is also robust in the sense that a single value of  $v_i$  affects the interpolated value much less than in quadrilinear interpolation or when using B-splines<sup>8</sup>. This makes the method very useful with respect to the problem of non-convexities and local optima. We implement a modified version which does not use all (millions) of grid points to compute the interpolated value because that would be infeasible, and we use only the closest 4 points on the grid in every dimension, with the weight on other points being zero. In the modified version the derivatives are not continuous but the discontinuities are less severe than in the case of quadrilinear interpolation which uses 16 points only (as opposed to 256). Other advantages of Shepard's interpolation relative to B-splines are that it is much easier to implement (particularly when it comes to parallel VFI), and the interpolated value can never be greater than the greatest of  $v_i$ 's, which is important for stability (as discussed by Judd (1998)).

To summarise, the modified Shepard's interpolation retains the best properties of quadrilinear interpolation (stability, ease of implementation), adds a near-continuity of first and second derivatives (which improves performance of local optimisation methods), and is

<sup>8</sup>The robustness of Shepard's interpolation is not caused only by the fact that a relatively great number of values is used in computing a continuation value, so the effect of every single one is smaller. B-splines of higher order than linear are particularly adversely affected by finding a 'very bad' solution at a point in state space, which usually happens at least at some points in the state space, when there are millions of points on the grid, and the feasible sets are not convex. Due to preserving continuity of derivatives, this outlier then usually positively affects the interpolated value at some neighbouring points, and this makes the region of the state space appear particularly attractive in terms of continuation value even if it is not, and the optimal choices in other points in state space tend to be biased in direction of this bad solution. In contrast, a bad solution at a point in state space affects Shepard's interpolation only in the sense that a close neighbourhood of the grid point where the bad solution occurred is less attractive the next iteration, and the algorithm then tends to avoid this problematic region. With sufficient exploration in future iterations we can eventually find a good solution even at the 'outlier' grid point.



relatively robust to errors in values used for interpolation, which is very useful when dealing with non-convex feasible sets.

## Deterministic steady-state

Obtaining a steady state in which all assets are kept at their initial level is useful for calibration, and for obtaining the initial guess for the first-period ( $t = 0$ ) problem in simulation. We need to solve a non-linear system of 4 equations (steady state version of the budget constraints of both governments, the resource constraint, and the net foreign debt evolution of country 1), where the unknowns are  $(c_{ss}^1, c_{ss}^2, l_{ss}^1, l_{ss}^2)$ :

$$\pi_1(g^1 - \theta_1 l_{ss}^1 + B(c_{ss}^1)^\sigma (l_{ss}^1)^{1+\gamma}) - \frac{\beta - 1}{\beta} b_{-1}^1 = 0 \quad (\text{C.14})$$

$$\pi_2(g^2 - \theta_1 l_{ss}^2 + B(c_{ss}^2)^\sigma (l_{ss}^2)^{1+\gamma}) - \frac{\beta - 1}{\beta} b_{-1}^2 = 0 \quad (\text{C.15})$$

$$\pi_1(c_{ss}^1 + g^1 - \theta_1 l_{ss}^1) + \pi_2(c_{ss}^2 + g^2 - \theta_2 l_{ss}^2) = 0 \quad (\text{C.16})$$

$$\pi_1(c_{ss}^1 + g^1 - \theta_1 l_{ss}^1) - \frac{\beta - 1}{\beta} b_{1,-1}^{ex} = 0 \quad (\text{C.17})$$

As for obtaining an initial guess for each point in state space before starting value function iteration, it is more appropriate to use an autarky steady state for each country, which does not take the initial asset levels into account<sup>9</sup>. In this case we obtain a much simpler system of equations for each of the countries (one implementability constraint and one resource constraint). If the utility is logarithmic in consumption ( $\sigma = 1$ ) and the initial debt is zero, then labour supply in country  $i$  given as

$$l^i = \left( \frac{1}{B} \right)^{1/(1+\gamma)}, \quad (\text{C.18})$$

and consumption is given by the autarky resource constraint  $c^i = \theta_i l^i - g_i$ .

---

<sup>9</sup>The reason is that in some parts of the state space the asset levels may be relatively far from the initial conditions (even opposite sign), and using as an initial guess a choice for such steady state is worse than not taking the asset levels into account at all.

